

Get Started with Tableau Prep Builder

Version: Current

Note: Starting in version 2020.4.1, as a Creator, you can also create and edit flows on the web. This tutorial was designed using Tableau Prep Builder, but can also be done on the web with some noted exceptions.

This tutorial introduces you to the common operations that are available in Tableau Prep. Using the sample data sets that come with Tableau Prep, you will walk through creating a flow for Sample Superstore. This tutorial uses the most current version of Tableau Prep Builder. If you are using a previous version, your results may differ.

Watch for tips along the way to gain insights into how Tableau Prep helps you clean and shape your data for analysis.

To install Tableau Prep Builder before continuing with this tutorial, see [Install Tableau Desktop or Tableau Prep Builder from the User Interface](#) in the Tableau Desktop and Tableau Prep Builder Deployment guide. Otherwise you can download the [free trial](#).

Sample files

To complete the tasks in this tutorial, you need to install Tableau Prep Builder, or if web authoring is enabled on your server version 2020.4 or later, you can also try the steps on the web.

After installing Tableau Prep Builder on your machine, you can also find the sample files in the following location :

- (Windows) C:\Program Files\Tableau\Tableau Prep Builder <version>\help\Samples\en_US\Superstore Files
- (Mac) /Applications/Tableau Prep Builder <version>.app/Contents/help/Samples/en_US/Superstore Files

Alternatively, download the sample files from these links and create a Samples directory and a South sub-directory. You'll need to do this if completing this tutorial on the web.

DOWNLOAD TO SAMPLES
DIRECTORY

- Orders Central
(https://mkt.tableau.com/PrepTutorials/Orders_Central.csv?fsi=IZz3CIwk).
- Orders East
(https://mkt.tableau.com/PrepTutorials/Orders_East.xlsx?fsi=IZz3CIwk)
- Orders West
(https://mkt.tableau.com/PrepTutorials/Orders_West.csv?fsi=IZz3CIwk).
- returns reasons new
(https://mkt.tableau.com/PrepTutorials/return_rea

DOWNLOAD TO SOUTH SUB-
DIRECTORY

- Orders South 2015
(https://mkt.tableau.com/PrepTutorials/orders_south_2015.csv?fsi=IZz3CIwk).
- Orders South 2016
(https://mkt.tableau.com/PrepTutorials/orders_south_2016.csv?fsi=IZz3CIwk).
- Orders South 2017
(https://mkt.tableau.com/PrepTutorials/orders_south_2017.csv?fsi=IZz3CIwk).

~~SOUTH.SAMPLES
fsi=IZz3CIwk)~~

DOWNLOAD TO SOUTH SUB-
DIRECTORY

- Orders_South_2018
([https://mkt.tableau.com/
PrepTutorials/orders_sou
th_2018.csv?
fsi=IZz3CIwk](https://mkt.tableau.com/PrepTutorials/orders_south_2018.csv?fsi=IZz3CIwk)).

Here's the story...

You work at the headquarters for a large retail chain. Your boss wants to analyze product sales and profits over the last four years for the company. You suggest that he use Tableau Desktop to do that. Your boss thinks that's a great idea and wants you to get right on that.

As you start gathering all the data you'll need, you notice that the data has been collected and tracked differently for each region. You also notice a lot of creative data entry in the different files, and that one region even has a separate file for each year!

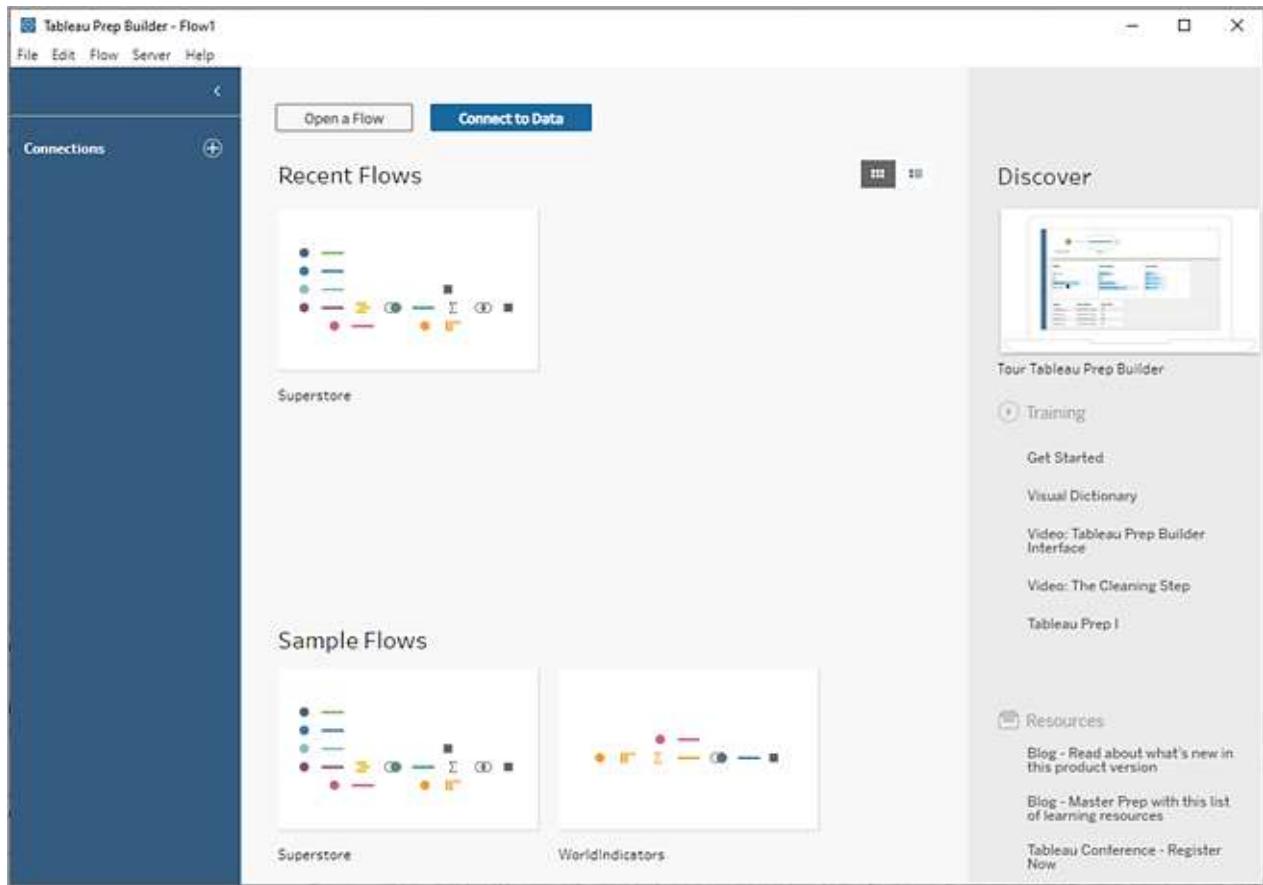
Before you can start analyzing the data in Tableau, you'll have to do some serious data cleaning first, and it's going to be a long night.

As you rummage for restaurant menus to order some dinner, you remember that Tableau has a product called Tableau Prep that might help you with your Herculean data cleaning tasks.

You download the product, or sign up for a [free trial](#) and decide to give it a try.

1. Connect to data

The first thing you see when you open Tableau Prep Builder is a Start page with a **Connections** pane, just like Tableau Desktop.



To get started, the first step is to connect to your data and create an Input step. From there you will start building a workflow or "flow", as it's called in Tableau Prep, and add more steps to take action on your data as you go.

Tip: The Input step is the ingestion point for your data and the starting point for your flow. You can have multiple Input steps and some might include multiple data files. For more information about connecting to data, see [Connect to Data](#).

Your sales data files for the different regions are stored in different formats, and your orders from the South are actually multiple files. You check out the **Connections** pane and see that you have a lot of choices to connect to data. Great!

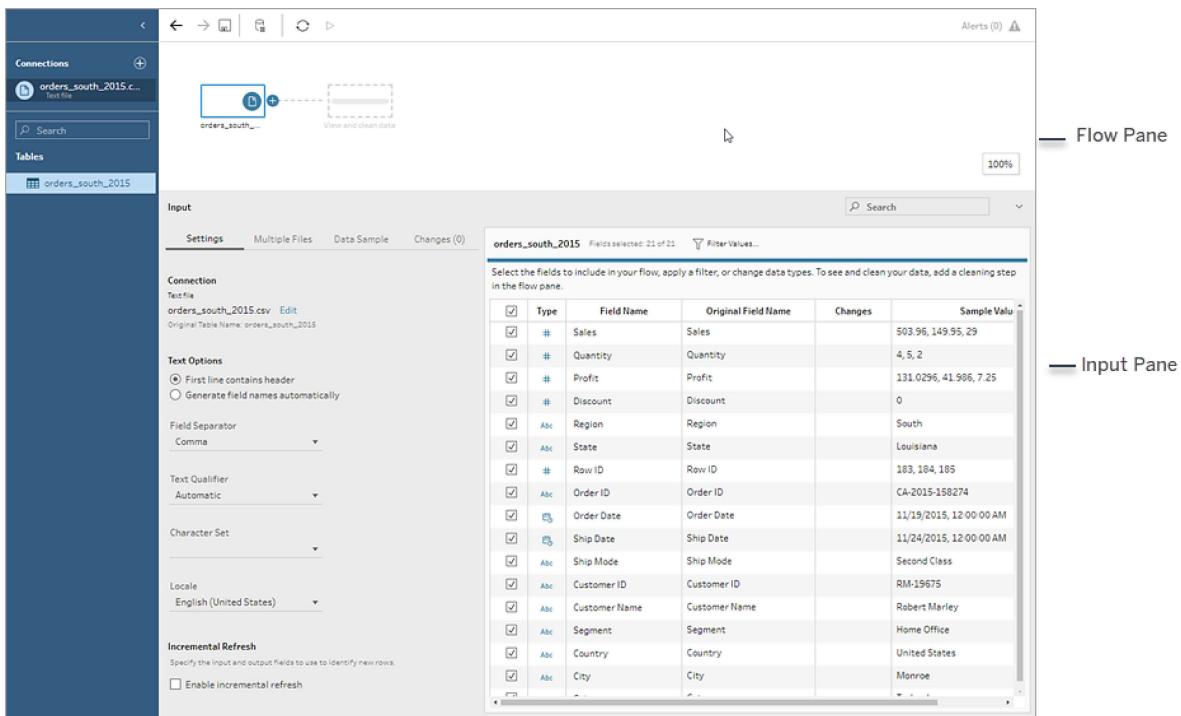
Since your other regions have one file for all four years worth of data, you decide to tackle the files from the South first.

1. On the **Connections** pane, click the **Add connection**  button.

In web authoring, from the **Home** page, click **Create > Flow** or from the **Explore** page, click **New > Flow**. Then click **Connect to Data**.

2. The files are .csv files, so select **Text file** in the list of connections.
3. Navigate to the directory for your files. In the Orders South subdirectory, select the first file **orders_south_2015.csv** and click **Open** to add it to your flow. (For file location, see [Wrap up and resources](#).)

After you connect to your first file, the Tableau Prep Builder workspace opens and you see it is divided into two main sections. The **Flow** pane at the top and the **Input** pane at the bottom.



Much like Tableau Desktop, this **Flow** pane is your workspace, where you can interact with your data visually and build your flow. The **Input** pane contains configuration options about how the data is ingested. It also shows you the fields, data types, and examples of your values from your data set.

We'll look at how you can interact with this data in the next section.

Tip: For single tables, Tableau Prep automatically creates an Input step for you in the **Flow** pane when you add data to your flow. Otherwise you can use drag-and-drop to add tables to the **Flow** pane.

4. You have three other files for your orders in the South, and how you combine them depends on where you're working.

In Tableau Prep Builder:

- a. You could add each file individually, but you want to combine all the files together into one Input step, so you click the **Multiple Files** tab in the **Input** pane.
- b. You see an option for **Wildcard union**. Select it.

You notice that the directory where you selected your file is already populated and the other files you need are listed in the **Included files** section in the Input pane.

Tip: Using a wildcard union is a great way to connect to and combine multiple files from a single data source with a similar name and structure. To use this option, the files must be in the same parent or child directory. If you don't see the files you need right away, change your search criteria. For more information, see [Union files and database tables in the Input step](#).

#	Type	Field Name	Original Field Name	Changes	Sample Value
1	Text	Date	Date	01/01/1990, 12/31/2018	2019-06-14, 95, 28
2	Text	Quantity	Quantity		4,5, 8
3	Text	Freight	Freight		131.0296, 41.906, 7.25
4	Text	Discount	Discount		2
5	Text	Region	Region		South
6	Text	State	State		Unassigned
7	Text	Row ID	Row ID		123,124,125
8	Text	Order ID	Order ID		CA-2018-12345678
9	Text	Order Date	Order Date		21/12/2018, 12:00:00 AM
10	Text	Ship Date	Ship Date		21/12/2018, 12:00:00 AM
11	Text	Ship Mode	Ship Mode		Second Class
12	Text	Customer ID	Customer ID		BA12345678
13	Text	Customer Name	Customer Name		Roland Marley
14	Text	Segment	Segment		Home Office
15	Text	Country	Country		United States
16	Text	City	City		Mountain View
17	Text	Postal Code	Postal Code		94035
18	Text	Product ID	Product ID		TEC-PA-10000273, TEC-PA-10000489, TEC-AC-10002345
19	Text	Category	Category		Technology
20	Text	Sub-Category	Sub-Category		Peripherals, Accessories
21	Text	Product Name	Product Name		ASUS F2230H, Nokia Lumia 521 (White), HP Standard 128 Key PS-2 Key

- c. Click **Apply** to add the data from these files to the **orders_south_2015** input step.

- d. The files for the other regions are all single table files, so you can select all of the files at once and add them to your flow.

Note: On the web, files can only be uploaded individually.

In Tableau Server or Tableau Online:

The wildcard option isn't currently available for Tableau Server or Tableau Online. Still, you want to include all of the files from the South and handle the data alike, so combining them makes sense.

- a. Repeat steps 2 and 3 to add the rest of the files from the Orders South subdirectory.
- b. Combine them with a union step. (For more details, see [Union files and database tables in the Input step](#) .)
 - i. Drag **Orders_South_2016** on top of **Orders_South_2015** and drop it on the **Union** option.



- ii. Drag **Orders_South_2017** on top of the new Union step and drop it on **Add**. Repeat this step with the final file.



5. Add the remaining files.

In Tableau Prep Builder:

- Open File Explorer or Finder and navigate to the directory for the files. Ctrl-click or Cmd-click (MacOS) to select the following files and drag-and-drop them onto the **Flow** pane to add them to your flow. (For file location, see [Wrap up and resources](#).)
 - Orders_Central.csv
 - Orders_East.xlsx
 - Orders_West.csv

Note: These are different file types. If you don't see all of these files, make sure your file explorer or finder is set to view all file types.

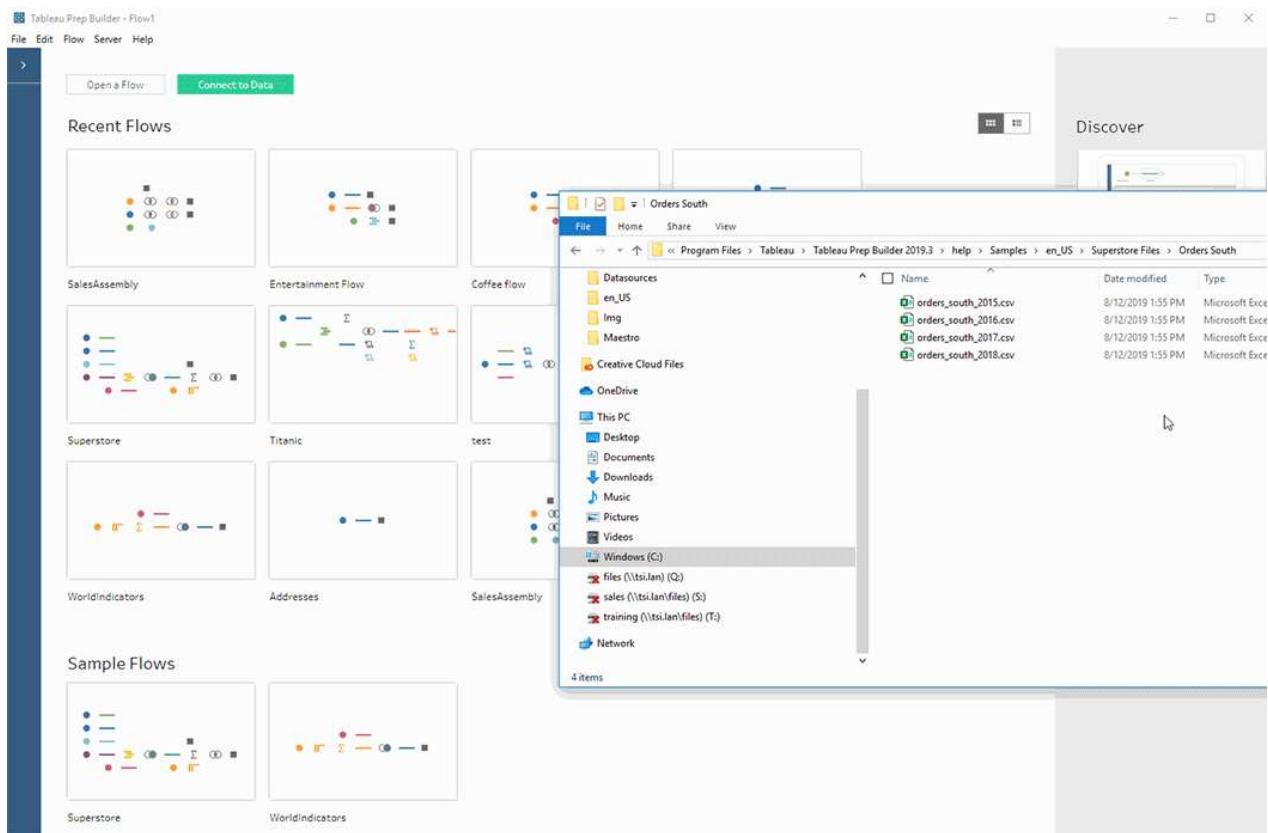
The screenshot shows the Tableau Data Importer interface. On the left, the 'Connections' pane lists several CSV files: 'orders_east_2020.csv', 'Orders_Middle...', 'Orders_Central...', and 'Orders_East.xlsx'. The 'Orders_East.xlsx' file is selected. The main area is titled 'Input' and shows the 'Multiple File' tab selected. A table preview for 'orders_east_2020.csv' is displayed, listing fields like 'Order ID', 'Customer ID', 'Customer Name', 'Segment', 'Country', 'City', 'Postal Code', 'Product ID', and 'Category'. The 'Segment' field is highlighted with a yellow background. The bottom right corner of the preview table contains the text 'Selected 1000'.

In Tableau Server or Tableau Online:

- Follow steps 2 and 3 to add Orders_Central.csv and Orders_West.csv.
- On the **Connections** pane, click the **Add connection** button. Click **Microsoft Excel** and select **Orders_East.xlsx**.

Check your work: Watch "Connect to data" in action.

Click the image to replay it



2. Explore your data

Now that you have the data files loaded into Tableau Prep, you're pretty sure that you want to combine all of the files together. But before you do that, it might be a good idea to take a look at them first and see if you can spot any issues.

When you select an Input step in the **Flow** pane, you can see the settings used to bring in the data, the fields that are included, and a preview of your values.

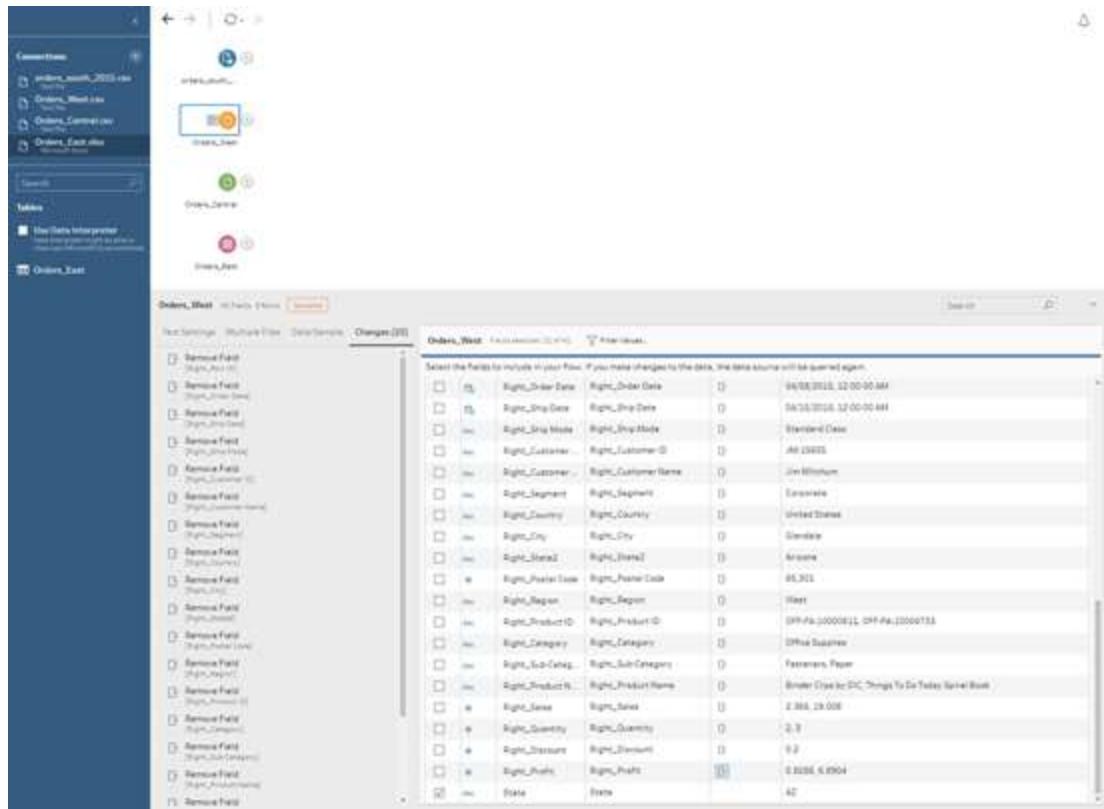
This is a good place to decide how much data you want to include in your flow and remove or filter fields that you don't want. You can also change any data types that were assigned incorrectly.

Tip: If you are working with large data sets, Tableau Prep automatically brings in a sample of the data to maximize performance. If you don't see the data you expect, you might need to adjust the sample. You can do this on the **Data Sample** tab. For more information about configuring your data options and sample size, see [Set your data sample size](#).

In the **Flow** pane, as you select each step and look over each data set, you notice a few things that you want to fix later and one thing that you can fix now in the **Input** step.

- Select the **Orders_West** Input step.
 - The **State** field uses abbreviations for the state name. Other files spell this out, so you'll need to fix that later.
 - There are a lot of fields that start with **Right_**. These fields appear to be duplicates of the other fields. You don't want to include these duplicate fields in your flow. This is something you can fix right here in the **Input** step:

To fix this now, clear the check box for all fields that start with **Right_**. This tells Tableau Prep to ignore these fields and not to include them in the flow.



Tip: When you perform cleaning operations in a step, like removing fields, Tableau Prep tracks your changes in the **Changes** pane and adds an annotation (in the form of a little icon) in the Flow pane to help you keep track of the actions you take on your data. For Input steps, an annotation is also added to each field.

- In the **Flow** pane, click the **Orders_Central** Input step to select it. In the **Input** pane, you notice the following issues:

- The order dates and ship dates are separated out into fields for month, day, and year.
- Some of the fields have different data types than the same fields in other files.
- There is no field for **Region**.

You'll need to do some cleaning on these fields before you can combine this file with the others files. But you can't fix that here in the **Input** step, so you make a note to do this later

- Select the **Orders_East** Input step.

The fields in this file look like they align pretty well with the other files. But the **Sales** values all seem to have the currency code included. You'll need to fix that later, too.

Now that you've identified a few troublemakers in your data sets, the next step is to examine your data a bit more closely and clean up any issues that you find so that you can combine and shape your data and generate an output file that you can use for analysis.

3. Clean your data

In Tableau Prep, examining and cleaning your data is an iterative process. After you decide on the data set that you want to work with, the next step is to examine and take action on that data by applying various cleaning, shaping, and combining operations to it. You apply these operations by adding steps to your flow. For more information about cleaning options, see [Clean and Shape Data](#)

Steps come in many flavors, depending on what you are trying to do. For example, add a cleaning step any time you want to apply cleaning operations to your fields like filter, merge, split, rename, and so on. Add an aggregation step to group and aggregate fields and change the level of detail of your data. For more information about the different step types and their uses, see [Build and Organize your Flow](#) .

Tip: As you add steps to your flow, a flow line is automatically added to connect the steps to one another. You can move these flow lines around and remove or add them as needed.

When you run your flow, these connection points are required so Tableau Prep knows which steps are connected and in which order the steps apply in the flow. If a flow line is missing, the flow will be broken and you'll get an error.

Clean Orders_Central

To address the issues you noticed earlier and to see if there are any other issues, you start by adding a cleaning step to the **Orders_Central** Input step.

1. In the **Flow** pane, select **Orders_Central**, do one of the following:
 - Click the plus  icon and add a cleaning step.
Depending on your version, this menu option is **Add Step**, **Add Clean Step**, or **Clean Step**.
 - Click on the suggested clean step (Tableau Prep Builder version 2020.3.3 and later and on the web)



When you add a cleaning step to your flow, the workspace changes and you see the details of your data.

A detailed screenshot of the Power BI workspace illustrating the three-part layout after adding a cleaning step:

- A. Flow pane:** On the left, showing the data flow from 'orders_south_2015.csv' through 'Orders_West.csv' and 'Orders_Central.csv' to 'Orders_East.xlsx', with a 'Clean 1' step in the middle.
- B. Toolbar:** A horizontal toolbar at the top of the main workspace area, featuring various icons for data analysis and visualization.
- C. Profile pane:** A vertical pane on the right side containing a toolbar, a dropdown menu for 'Segment', and a list of names categorized under 'Consumer', 'Corporate', and 'Home Office'.
- D. Data grid:** The main workspace area displaying a table with 24 fields and 2K rows. The table includes columns for Row ID, Order ID, Ship Mode, Customer ID, Customer Name, Segment, Country, City, State, Postal Code, Product ID, and Category. The 'Ship Mode' column shows categories like First Class, Same Day, Second Class, and Standard Class. The 'Customer Name' column lists names such as Aaron Bergman, Aaron Smayling, Adam Bellavance, Adam Hart, Adam Shillingburg, Adrian Barton, Adrian Hane, Aimee Bibby, Alan Barnes, Alan Dominguez, Alan Haines, and Alan Huang.

A. Flow pane, B. Toolbar, C. Profile pane, D. Data grid

The workspace is now split into three parts: the **Flow** pane, the **Profile** pane with a toolbar, and the **Data** grid.

The **Profile** pane shows you the structure of your data, summarizing the field values into bins so that you can quickly see related values and spot outliers and null values. The **Data** grid shows you the row level detail for your fields.

Tip: Each field in the **Profile** pane is shown on a profile card. Use the **More options...** menu (drop-down arrow in prior versions) on each card to see and select the different cleaning options that are available for that field type. You can also sort the field values, change the data type, assign a data role to the field or drag and drop the profile cards and the columns in the **Data** grid to rearrange them.

Clean data with calculated fields

This data set is missing a field for **Region**. Since the other data sets have this field you'll need to add it so that you can combine your data later. You'll need to use a calculated field to do this.

2. In the toolbar, click **Create Calculated Field**.

3. In the Calculation editor, name the calculated field **Region**. Then enter "**Central**" (including the quotes) and click **Save**.



You love the flexibility of being able to use calculated fields to shape your data. You are pleased to see that Tableau Prep uses the same calculation editor language as Tableau Desktop.

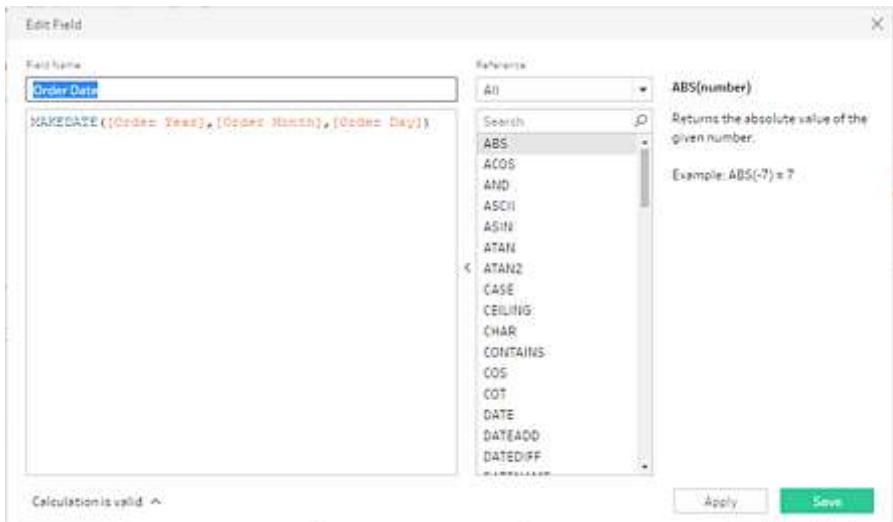
Tip: When you make changes to your fields and values, Tableau Prep keeps track of them in the **Changes** pane on the left. An icon (annotation) representing the change is also added to the cleaning step in the flow and to the field in the **Profile** pane. We'll look at the **Changes** pane after making more changes.

Next you want to address the separate order date and ship date fields. You want to combine them into two single fields, one for **Order Date** and one for **Ship Date** so they align with the same fields in the other data sets. Making sure your tables have the same fields will enable you to combine the tables using a union later.

You can use a calculated field again to do this in one easy step.

4. In the toolbar, click **Create Calculated Field** to combine the **Order Year**, **Order Month**, and **Order Day** fields into one field with the format "MM/DD/YYYY".
5. In the Calculation editor, name the calculated field **Order Date**. Then enter the following calculation and click **Save**:

```
MAKEDATE( [Order Year] , [Order Month] , [Order Day] )
```



Now that you have a new field for your order date, you want to remove the existing fields, as you no longer need them.

You have a lot of fields in the **Profile** pane. You notice a **Search** box in the top right corner on the toolbar. You wonder if you can use that to quickly find the fields that you want to remove. You decide to give it a try.

6. In the **Profile** pane, in the search box, type **Order**.

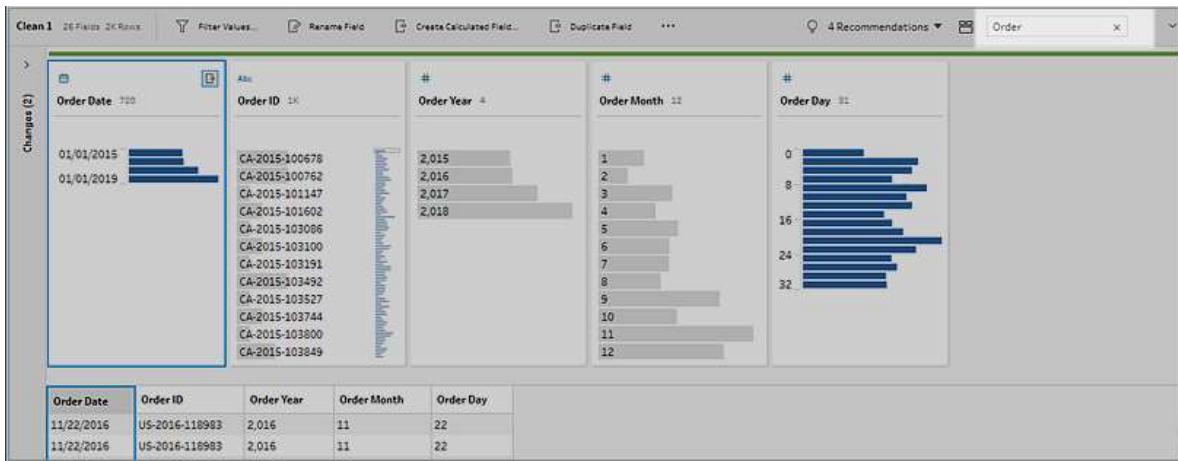


Tableau Prep quickly scrolls all the fields with **Order** in the name into view. Cool!

7. Ctrl-click or Cmd-click (MacOS) to select the fields for **Order Year**, **Order Month**, and **Order Day**. Then right-click on the selected fields and select **Remove** (**Remove Field** in prior versions) from the menu to remove them.



8. Now repeat steps 4 though 7 above to create a single field for **Ship Date**. Try it on your own or use the steps below to help you.

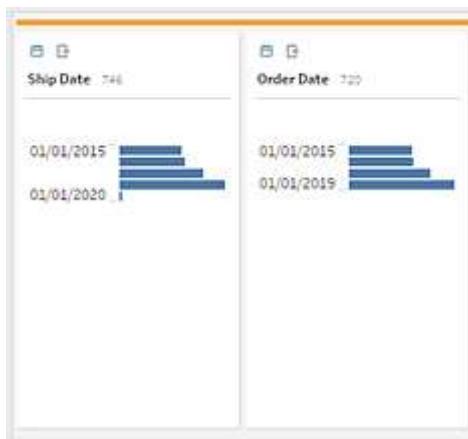
- In the toolbar, click **Create Calculated Field** to combine the **Ship Year**, **Ship Month**, and **Ship Day** fields into one field with the format "MM/DD/YYYY".
- Name the calculated field **Ship Date** and enter the calculation `MAKEDATE([Ship Year], [Ship Month], [Ship Day])`. Then click **Save**.
- Remove the **Ship Year**, **Ship Month**, and **Ship Day** fields. Search for the fields, select them, and select **Remove (Remove Field in prior versions)** from the menu to remove the fields.

Tip: Tableau Prep summarizes the data in the Profile pane into bins to help you quickly see the shape of your data, find outliers, spot relationships between fields, and so on.

In this scenario, the order and ship dates can now be summarized by year. Each bin represents a year from

January of the starting year to January of the following year and is labeled accordingly. Because there are sales dates and ship dates that fall in the latter part of 2018 and 2019, we get a bin for that data that is labeled with the ending year 2019 and 2020 accordingly.

To change this view to the actual dates, click the **More options** ... menu (drop-down arrow in prior versions) in the Profile card and select **Detail**.



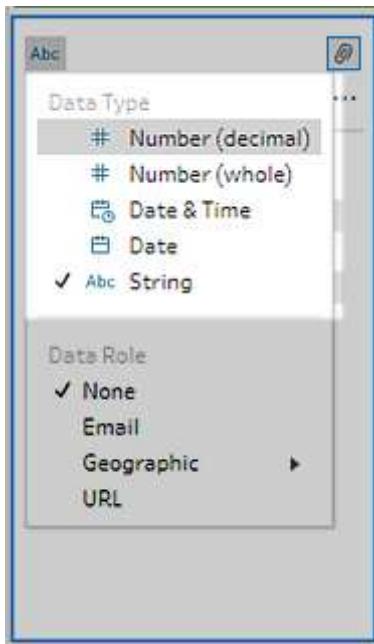
Interact directly with fields to clean your data

Your data is starting to look good. But, as you finish removing the extra fields for the order and ship dates, you notice that the **Discounts** field has a couple of issues.

- It's assigned to a **String** data type instead of a **Number (decimal)** data type.
- There's a field value **None** instead of a numeric value for no discount.

This will cause a problem when you combine the files, so you better fix that too.

9. Clear your search and enter **disc** in the search box to find the field.
10. Select the **Discounts** field, double-click the field value **None**, and change it to the numeric value **0**.
11. To change the data type for the **Discount** field from **String** to **Number (decimal)**, click **Abc** and select **Number (decimal)** from the drop-down menu.



12. Finally name your step to help keep track of what you did in this step. In the **Flow** pane, double-click the step name **Clean 1** and type in **Fix dates/field names**.

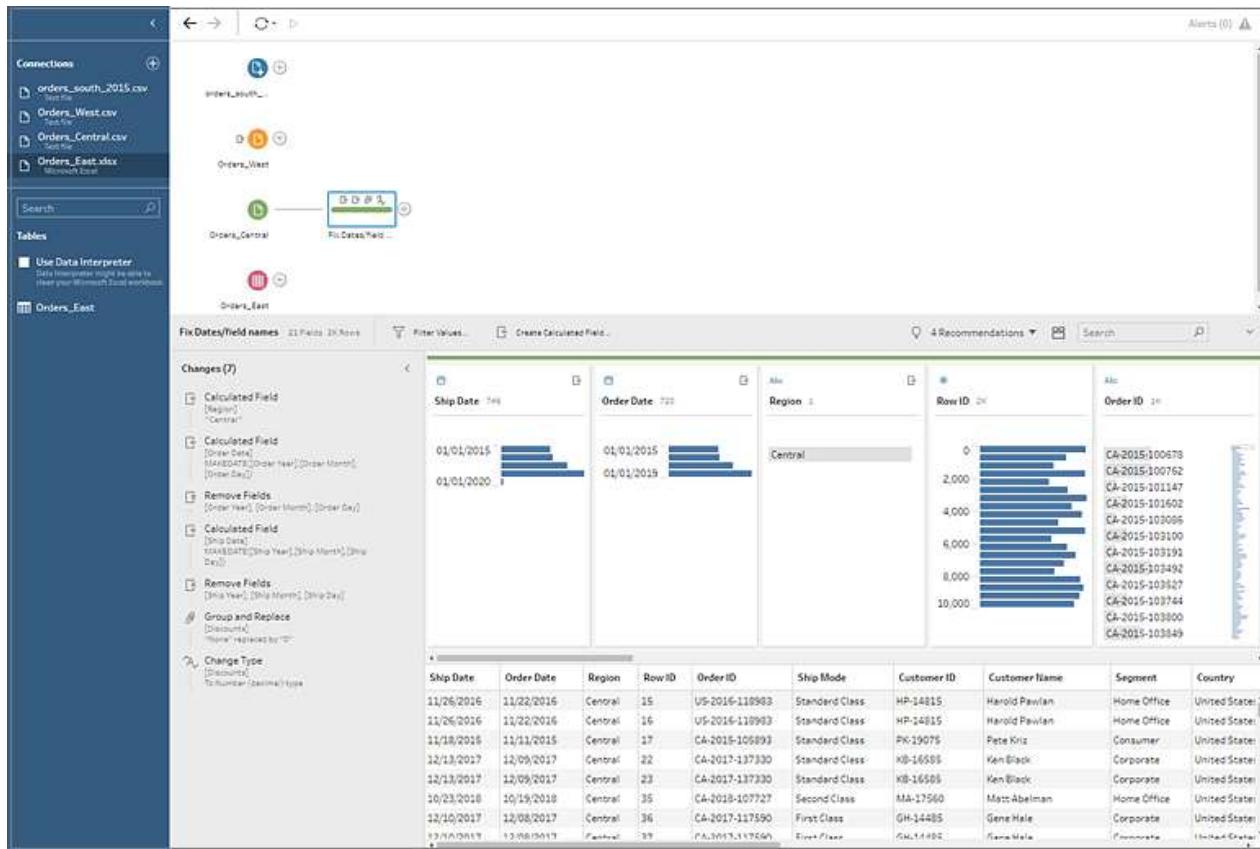
Review your changes

You made a lot of changes to this data set and you start to worry that you won't remember everything you did. As you look over your work, you see a column on the left of the **Profile** pane called **Changes**.

You click the arrow to open it and are delighted to see a list of every change you just made. As you scroll through the changes in the list, you notice that you can delete or edit your changes

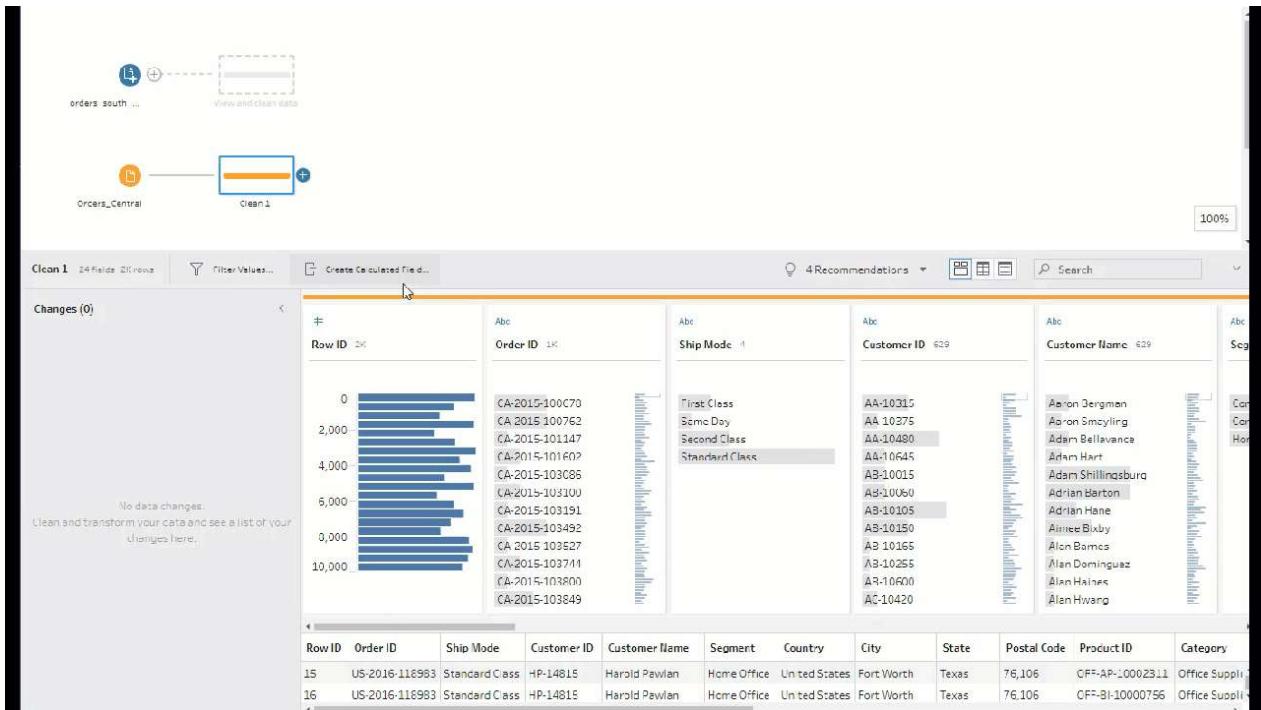
or even move them around to change the order that you did them in.

You love that you can easily find the changes you made in any step as you build your flow and experiment with the order of those changes to get the most out of your data.



Check your work: Watch "Clean Orders_Central" in action.

Click the image to replay it



Now that you've cleaned one file, you take a look at the other files to see what other issues you need to fix.

You decide to look at the Excel file for **Orders_East** next.

Clean Orders_East

As you look over the fields for the **Orders_East** file, most of the fields look like they align with the other files, except for **Sales**. To take a closer look and see if there are any other issues to address, you add a cleaning step to the **Orders_East** Input step.

1. In the **Flow** pane, select **Orders_East** and do one of the following:

- Click the plus  icon and add a clean step. Depending on your version, this menu option is **Add Step**, **Add Clean Step**, or **Clean Step**.
- Click on the suggested clean step (Tableau Prep Builder version 2020.3.3 and later and on the web).

Looking at the **Sales** field you quickly see that the **USD** currency code has been included with the sales numbers, and Tableau Prep interpreted these field values as a string.

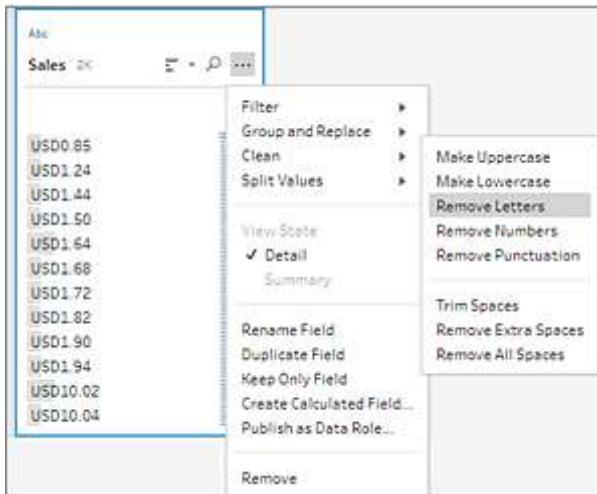
You'll need to remove the currency code from this field and change the data type if you want to get accurate sales data.

Fixing the data type is easy, you already know how to do that. But there are over 2000 unique rows of sales data and fixing every individual row to remove the currency code seems cumbersome.

But this is Tableau Prep, and you decide to check out the drop-down menu to see if there is an option to fix this.

When you click the **More options** ... (drop-down arrow in prior versions) for the **Sales** field, you see a menu option called **Clean** and an option under that to remove letters. You decide to give that a try and see what it does.

2. Select the **Sales** field. Click the **More options** ... menu (drop-down arrow in prior versions) and select **Clean > Remove Letters**.



Wow! That cleaning option instantly removed the currency code from every field. Now you just need to change the data type from **String** to **Number (decimal)** and this file is looking good.

3. Click the data type for the **Sales** field and select **Number (decimal)** from the drop-down list to change the data type.

4. The rest of the file looks pretty good. Name your cleaning step to keep track of your work. For example, **Change data type**.

Next you look at your last file for **Orders_West** to see if there are any issues there that you need to fix.

Clean Orders_West

As you look over the fields for the **Orders_West** file, most of the fields look like they align with the other files, but you remember seeing that the **States** field used abbreviations for

the values instead of spelling out the state name. To combine this file with the other files, you'll need to fix this. So you add a cleaning step to the **Orders_West** Input step.

1. In the **Flow** pane, select **Orders_West** and do one of the following:
 - Click the plus  icon and add a clean step.
 - Click on the suggested clean step (Tableau Prep Builder version 2020.3.3 and later and on the web).
2. Scroll or use Search to find the **State** field.

You see that all the state name values use the short abbreviation. There are only 11 unique values for this field. You could manually change each one, but maybe Tableau Prep has another way to do this?

You click the **More options ...** menu (drop-down arrow in prior releases) for the field and see an option called **Group Values** (Group and Replace in previous versions). When you select it you see several options:

- Manual Selection
- Pronunciation

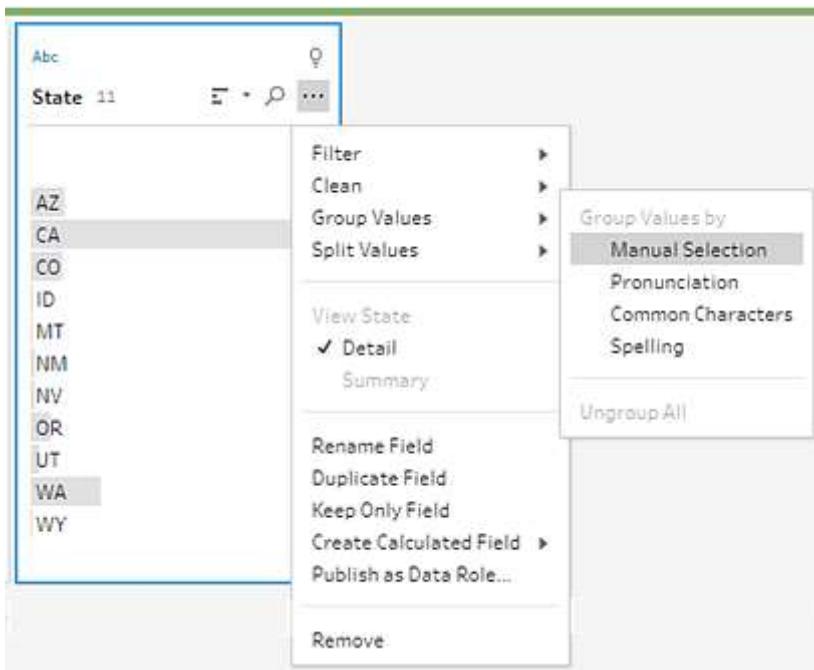
- Common Characters
- Spelling

The state names don't sound alike, they aren't spelled incorrectly, and they don't share the same characters, so you decide to try the **Manual Selection** option.

Tip: You can double-click a field name or field value to edit a single value. To edit multiple values you can select all the values and use the right-click menu option **Edit Values**. But when you want to map one or more values to specific values, use the **Group Values** option in the drop-down menu.

For more information about editing and grouping values, see [Edit field values](#) .

3. Select the **State** field. Click the drop-down arrow and select **Group Values** (Group and Replace in previous versions) > **Manual Selection**.



A two column card opens. This is the **Group Values editor**. The column on the left shows the current field values and the column on the right shows the fields that are available to map to the fields on the left.

You want to map your state abbreviations to the spelled out version of the state name, but you don't have those values in the **Orders_West** data set. You wonder if you can just edit the name directly and maybe add it there, so you give that a try.

4. In the **Group Values editor** in the left pane, double-click **AZ** to highlight the value and type **Arizona**. Then press **Enter** to add your change.

Group Values by Manual Selection		Done
State 11	AZ 1 member	
Arizona	<input checked="" type="checkbox"/> AZ	
CA	<input type="checkbox"/> CA	
CO	<input type="checkbox"/> CO	
ID	<input type="checkbox"/> ID	
MT	<input type="checkbox"/> MT	
NM	<input type="checkbox"/> NM	
NV	<input type="checkbox"/> NV	
OR	<input type="checkbox"/> OR	
UT	<input type="checkbox"/> UT	
WA	<input type="checkbox"/> WA	
WY	<input type="checkbox"/> WY	

Tableau Prep created a mapped value for your new value **Arizona** and automatically mapped the old value, **AZ** to it. Having a mapped relationship set up for these values will save you time if you get more data from this region entered like this.

Tip: You can add field values that aren't in your data sample to set up mapping relationships to organize your data. If you refresh your data source and new data is added, you can add the new data to the mapping instead of manually fixing each value.

When you manually add a value that isn't in your data sample, the value is marked with a red dot to help you easily identify it.

5. Repeat these steps to map each state to the spelled out version of its name.

Abbreviation	State Name
AZ	Arizona
CA	California
CO	Colorado
ID	Idaho
MT	Montana
NM	New Mexico
NV	Nevada
OR	Oregon
UT	Utah
WA	Washington

WY

Wyoming

Then click **Done** to close the **Group Values** editor.

The screenshot shows the Tableau Prep Data Flow interface. At the top, there's a flow diagram with four inputs: 'Orders_Central', 'Orders_West', 'Orders_East', and 'Clean 1'. The 'Clean 1' step has a 'Group Values' action applied to the [State] field. In the center, there's a 'Changes' pane showing one entry for 'Group Values [State]'. To the right, there's a 'Group Values by Manual Selection' pane where states like Arizona, California, Colorado, Idaho, Montana, Nevada, New Mexico, Oregon, Utah, Washington, and Wyoming are listed with checkboxes for mapping. At the bottom, a preview table shows three rows of order data with columns: Row ID, Order ID, Order Date, Ship Date, Ship Mode, Customer ID, Customer Name, Segment, Country, City, Postal Code, Region, and Product.

After all the states are mapped, you look at the **Changes** pane and see there is only one entry there instead of 11.

Tableau Prep grouped similar actions for a field together. You like that because it will make it easier to find changes you made to your data set later.

Fixing the **State** field values was the only change you needed to make here.

6. Name your cleaning step to keep track of your work. For example **Rename states**.

You've done a lot of clean up in your files, and you can't believe how quick and easy it was. You might make it home for dinner after all! To make sure that you don't lose all of your work so far, save your flow.

Note: If working on the web, your changes are automatically saved as you go, creating a draft flow. Click in the draft title to name your draft. For more information about authoring on the web, see **Tableau Prep on the Web** in the [Tableau Server](#) or [Tableau Online](#) help.

Click **File > Save** or **File > Save As**. Save your file as a flow file (.tfl) and give it a name. For example, **My Superstore**.

Tip: When you save your flow files, you can either save them as a flow file (.tfl) or you can save them as a packaged file (.tflx) and package your local data files with them to share the flow and files with someone else. For more information about saving and sharing your flows, see [Save and Share Your Work](#).

4. Combine your data

Now that all the files are cleaned up, you are finally ready to combine them all.

Because all the files have similar fields after your clean up efforts, to pull all the rows together into a single table, you need to union the tables.

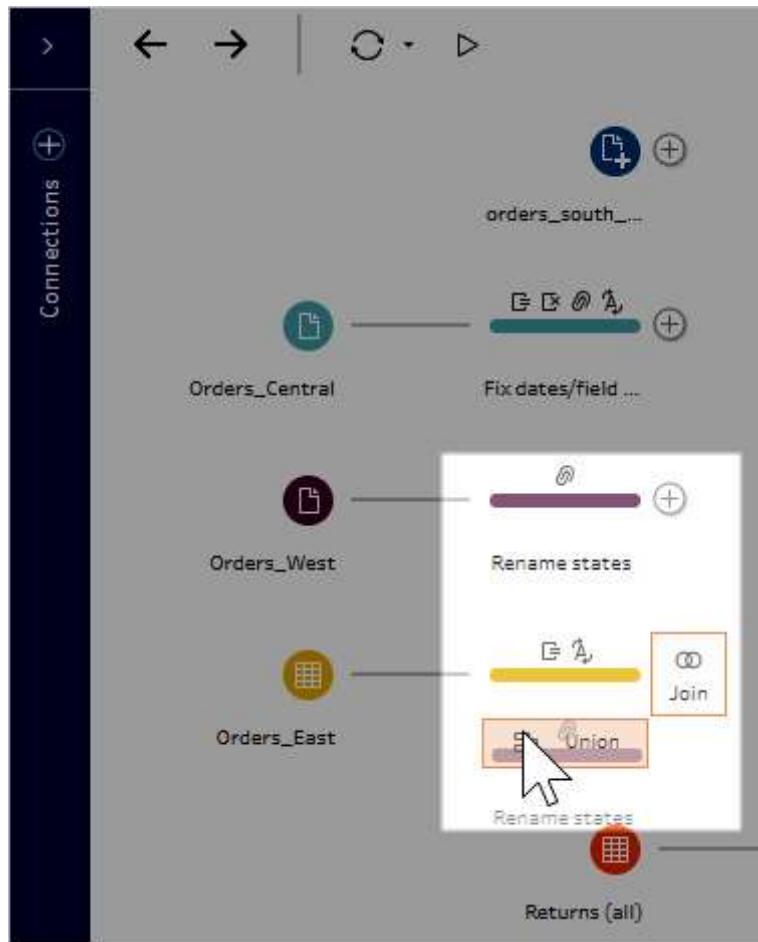
You remember that there was a step option called **Union**, but you wonder if you can simply drag and drop the steps to union them. You decide to try it and see.

Union your data

1. Follow the steps for where you are working.

Tableau Prep Builder

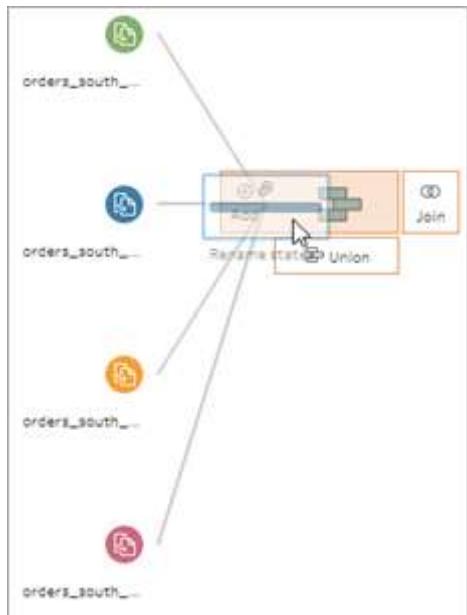
- In the **Flow** pane, drag the cleaning step **Rename states** on to the cleaning step **Changed data type** step and drop it on the **Union** option.



You see that Tableau Prep Builder added a new **Union** step to your flow. Great! Now you want to add the other files to this union too.

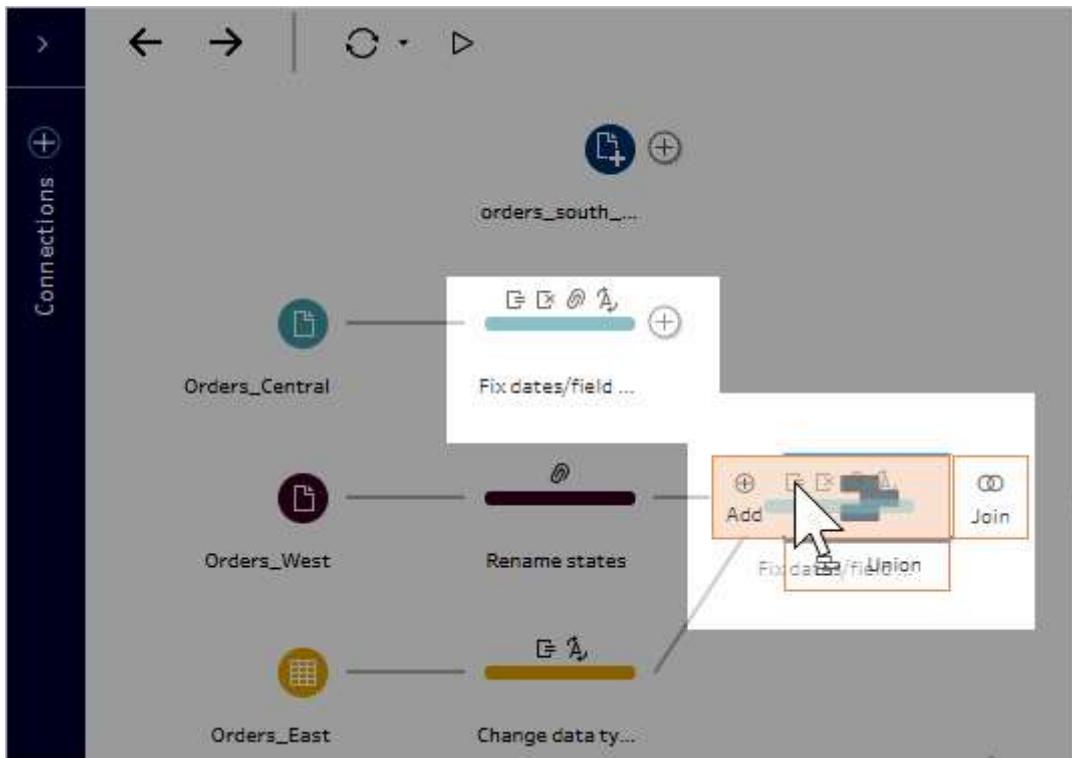
Tableau Server or Tableau Online

- In the **Flow** pane, drag the cleaning step **Rename states** on to the Union step you created earlier for your South files and drop it on the **Add** option.



You see that Tableau Prep added your new files to your previous union. Great! Now you want to add the other files to this union too.

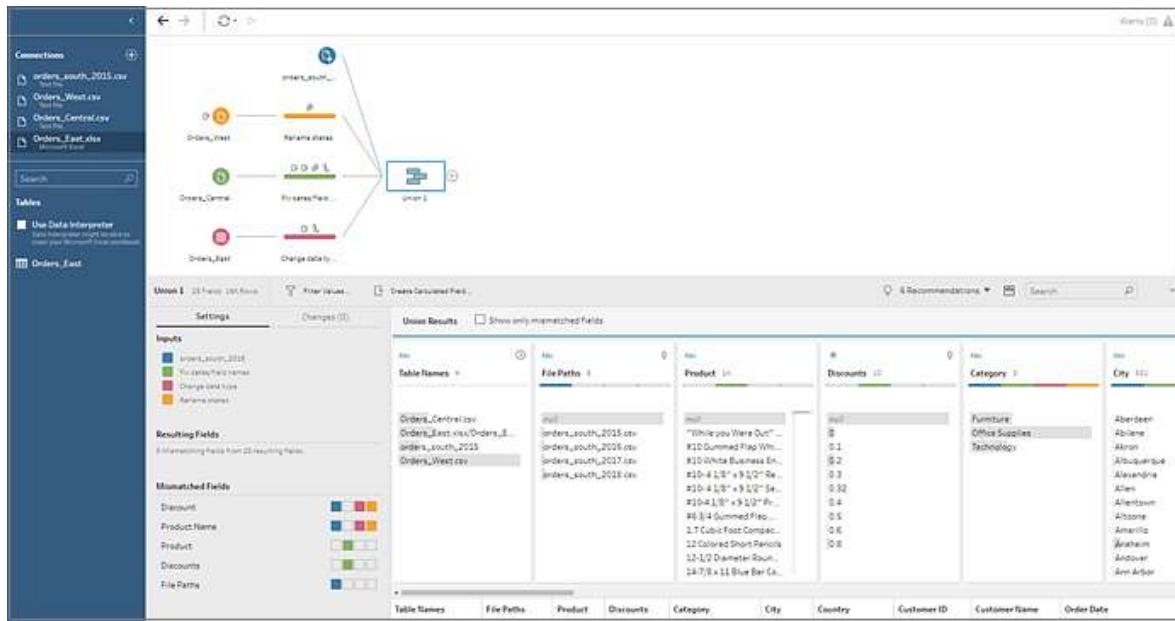
2. Drag the next cleaning step in the flow on to the **Union** step, then drop it on **Add** to add it to the existing union.



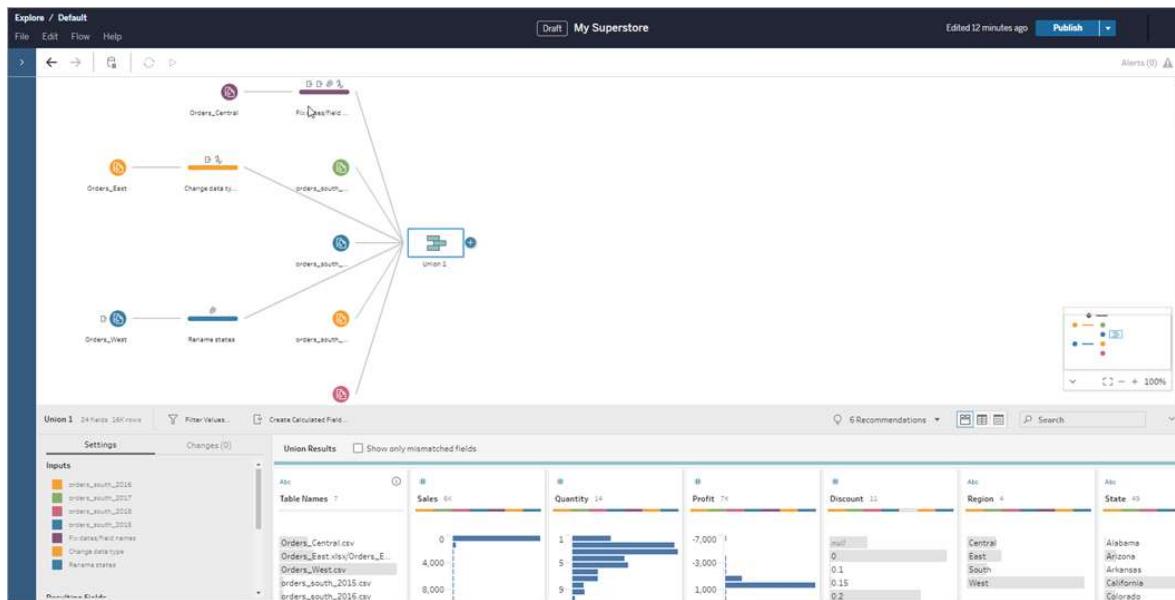
3. Drag the remaining step (**orders_south_2015** Input step if working in Tableau Prep Builder or your cleaning step if working on the web) to the new **Union** step. Drop it on **Add** to add it to the existing union.

Now all of your files are combined into a single table. In the **Flow** pane, select the new **Union** step to see your results.

On Tableau Prep Builder:



On Tableau Server or Tableau Online:



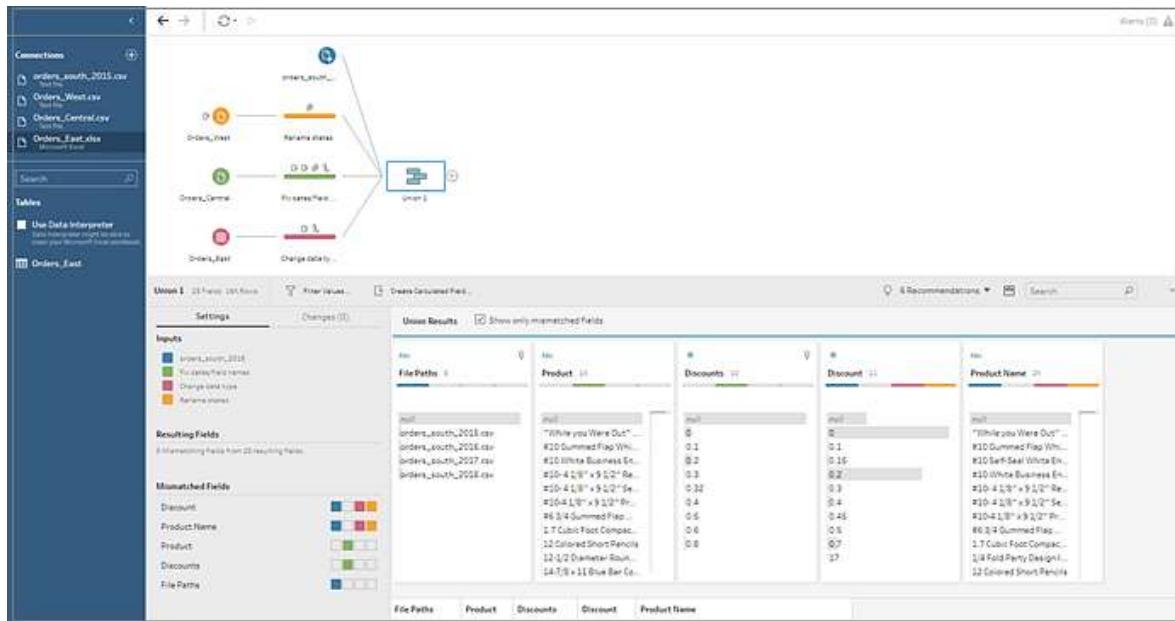
You notice that Tableau automatically matched up the fields that had the same names and types.

You also see that the **colors** assigned to the steps in the flow are used in the union profiles to indicate where the field came from and also appear in the **colored band** across the top of each field to show you if that field exists in that table.

You notice that a new field called **Table Names** was added that lists the tables where all the rows in the union come from.

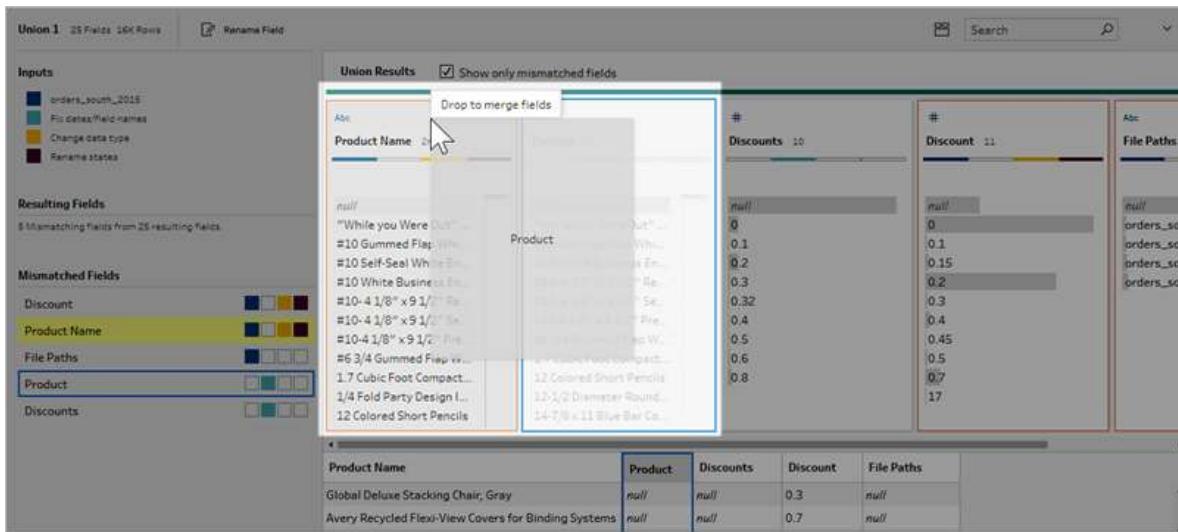
A list of mismatched fields also shows in the summary pane and you can see right away that the fields **Product** and **Discounts** only appear in the **Orders_Central** file.

4. To take a closer look at these fields, in the **Union Results** pane, select the **Show only mismatched fields** check box.



Looking at the field data, you quickly see that the data is the same, but the field name is different. You could simply rename the field, but you wonder if you could just drag and drop these fields to merge them. You decide to try that and see.

5. Select the **Product** field and drag and drop it onto the **Product Name** field to merge the fields. After the fields are merged, they no longer appear in the pane.



6. Repeat this step to merge the **Discounts** field with the **Discount** field.

The only field that doesn't have a match now is the **File Paths** field. In Tableau Prep Builder, this field shows the file paths for the wildcard union that you did for your sales orders from the South. You decide to leave this field there as it has good information.

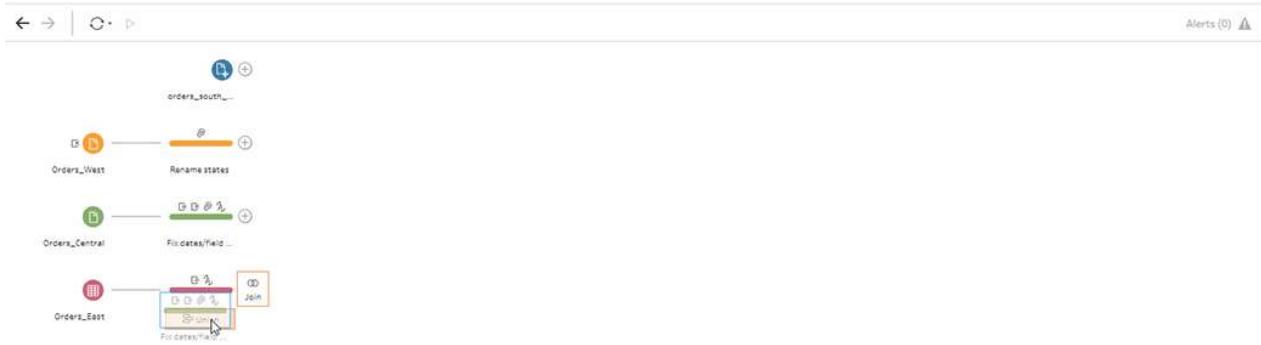
Tip: You have several options when fixing mismatched fields after a union. If Tableau Prep detects a possible match, it will highlight it in yellow. To merge the fields hover over the highlighted field and click the plus button that appears.

For more ways to merge fields in a union, see [Fix fields that don't match \(prep_combine.htm#Fix_union\)](#).

7. Clear the **Show only mismatched fields** check box to show all the fields included in the union.
8. Name your Union step to represent what this union includes work. For example, **All orders**.

Check your work: Watch "Union your data" in action.

Click the image to replay it



You are a cleaning genius! As you are admiring your results, your boss calls. He forgot to mention that he also wants you to include any product returns in your analysis. He hopes that won't be too much trouble. With Tableau Prep in your toolkit, it's no problem at all!

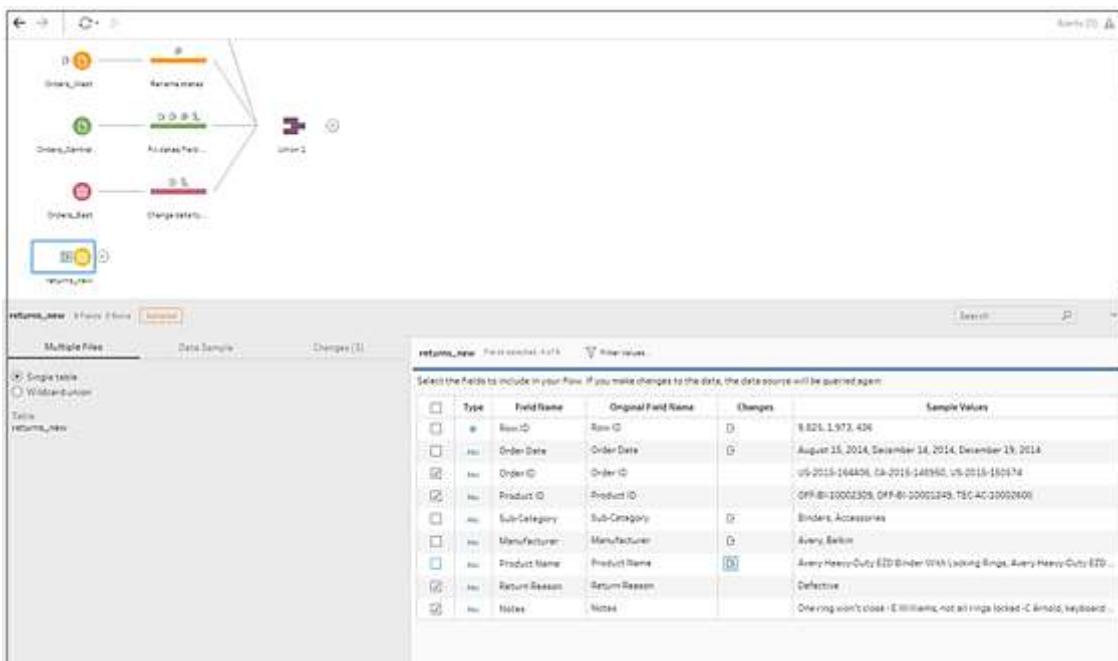
Clean the product returns data

You look over the Excel file that your boss sent you for product returns and it looks a little messy. You add the new file **return_reasons_new** to your flow to take a closer look.

1. In the **Connections** pane, click **Add connection**. Select **Microsoft Excel** and navigate to the sample data files you've been using for this exercise. (See [Sample files](#) to download the file.)
2. Select **return reasons_new.xlsx**, and then click **Open** to add the file to the flow pane.

There are only four fields that you want to include from this file in your flow: **Order ID**, **Product ID**, **Return Reason** and **Notes**.

3. In the **Input** pane for **returns_new** clear the check box at the top of the left-most column to clear all the check boxes. Then select the check box for the **Order ID**, **Product ID**, **Return Reason** and **Notes** fields.



4. Rename the Input step to better reflect the data that is included in this input. In the **Flow** pane, double-click the Input step name **Returns_new** and type in **Returns (all)**.

Looking at the sample field values, you notice that the **Notes** field seems to have a lot of different data combined together.

You have some cleaning to do in this file before you can do any further work with the data, so you add a cleaning step to check it out.

5. In the **Flow** pane, select the Input step **Returns (all)**, click the plus  icon or on the suggested clean step to add a clean step.

In the **Profile** pane, re-size the **Notes** field so you can see the entries better. To do this, click and drag the outer right edge of the field to the right.

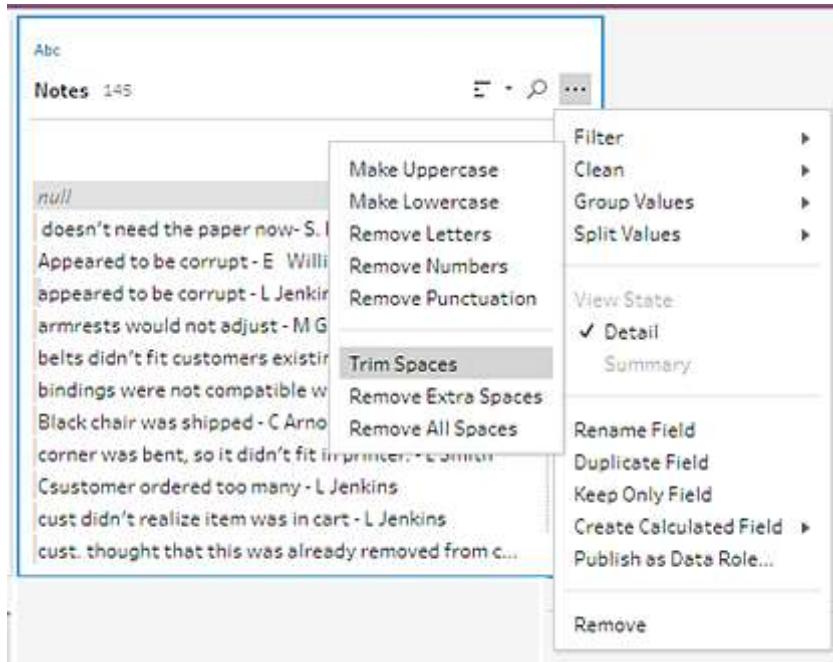
6. In the **Notes** field, use the visual scroll bar to the right of the field values to scan the values.

You notice a few things that are problematic:

- Some of the entries have an extra space in the entry.
This can result in the field being read as a null value.
- It looks like the name of the approver is included in the return notes entry. To better work with this data you'll want that information in a separate field.

To tackle the extra spaces, you remember that there was a cleaning option to remove trailing spaces, so you decide to try that to see if it can fix that problem.

7. Select the **Notes** field. Click the **More options...** menu (drop-down arrow in prior releases) and select **Clean > Trim Spaces.**



Yes! It did exactly what you wanted it to do. The extra spaces are gone.

Next you want to create a separate field for the approver name. You see a **Split Values** option in the menu, so you decide to try that.

8. Select the **Notes** field. Click the **More options...** menu (drop-down arrow in prior releases) and select **Split Values > Automatic Split.**

This option did exactly what you were hoping it would do. It automatically split the return notes and the approver name into separate fields.

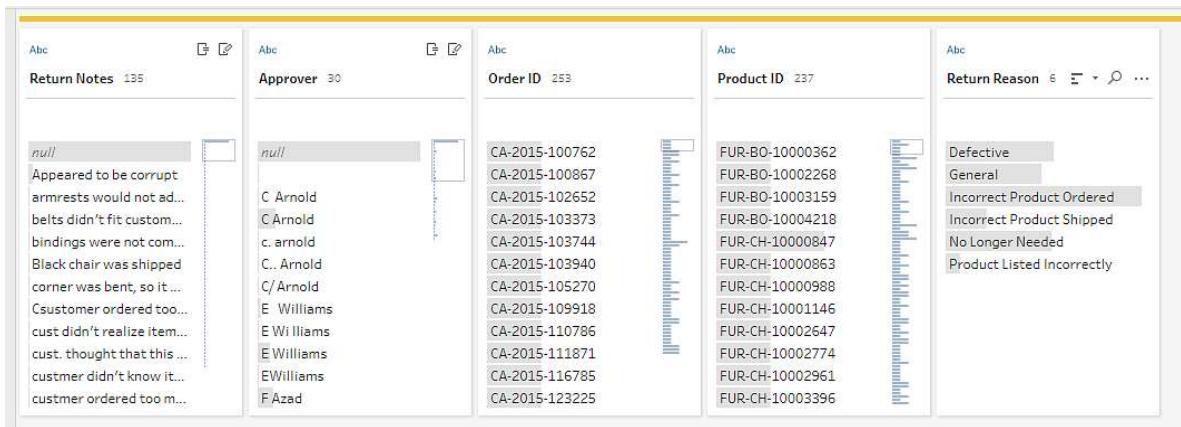
The screenshot shows two side-by-side columns in Tableau Prep. The left column is titled "Notes - Split 1" and contains 135 rows. The right column is titled "Notes - Split 2" and contains 30 rows. Both columns have a header row labeled "null".

Notes - Split 1	Notes - Split 2
null	
Appeared to be corrupt armrests would not ad...	C Arnold
belts didn't fit custom...	C Arnold
bindings were not com...	c. arnold
Black chair was shipped corner was bent, so it ...	C.. Arnold
Ccustomer ordered too...	C/ Arnold
cust didn't realize item...	E Williams
cust. thought that this ...	E Wi lliams
custmer didn't know it...	E Williams
custmer ordered too m...	EWilliams
	F Azad

Just like Tableau Desktop, Tableau Prep automatically assigned a name to those fields. So you'll need to rename the new fields to something meaningful.

9. Select the field **Notes-Split 1**. Double-click in the field name and type **Return Notes**.
10. Repeat this step for the second field and rename it to **Approver**.

11. Finally remove the original **Notes** field, as you no longer need it. Select the **Notes** field, click the **More options** ... menu (drop-down arrow in prior versions), and select **Remove (Remove Field** in prior versions) from the menu.

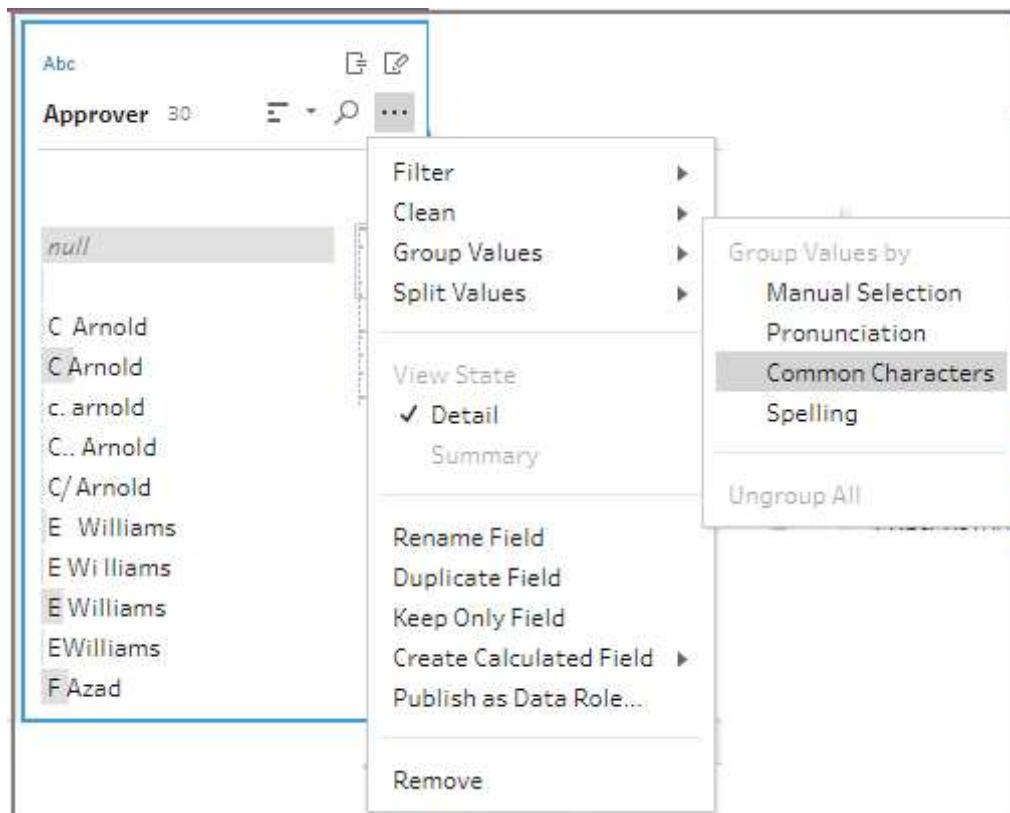


Looking at the new **Approver** field, you notice that the field values lists the same names but they are entered differently. You want to group them to eliminate multiple variations of the same value.

Maybe the **Group Values** (Group and Replace in prior versions) option can help with that?

You remember there was an option for **Common Characters**. Since these values share the same letters, you decide to try that.

12. Select the **Approver** field. Click the **More options...** menu (drop-down arrow in prior versions) and select **Group Values** (Group and Replace in prior versions) > **Common Characters**.



This option grouped all of the variations of each name together for you. That's exactly what you wanted to do.

After checking the other names to make sure they are grouped properly, you click **Done** to close the **Group Values editor**.

Group Values by Common Characters

Approver 13

C Arnold 5 members

Done

Approver	Grouped Value
null	
⑥ C Arnold	<input checked="" type="checkbox"/> C Arnold
⑥ E Williams	<input type="checkbox"/> E Williams
⑥ F Azad	<input type="checkbox"/> F Azad
⑥ G Lindsay	<input type="checkbox"/> G Lindsay
⑥ K Lawrence	<input type="checkbox"/> K Lawrence
⑥ L Jenkins	<input type="checkbox"/> L Jenkins
⑥ L Smith	<input type="checkbox"/> L Smith
⑥ M Gomez	<input type="checkbox"/> M Gomez
⑥ R Chen	<input type="checkbox"/> R Chen
⑥ R Duchesne	<input type="checkbox"/> R Duchesne

This file is looking pretty good.

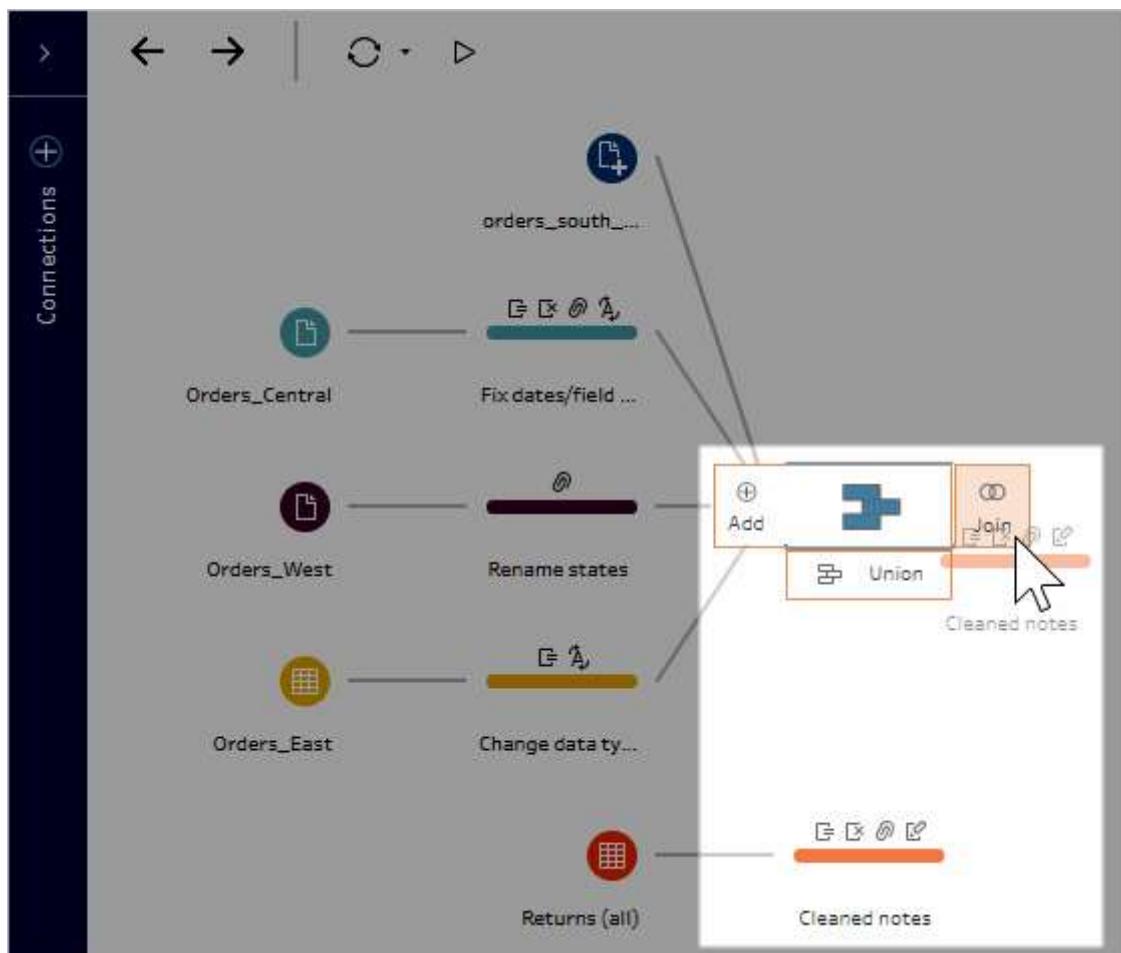
13. Name your cleaning step to keep track of your work. For example **Cleaned notes**.

Now that the product return data is all cleaned up, you want to add this data to the orders data from your unioned files. But many of these fields don't exist in the unioned files. To add these fields (columns of data) to your unioned data set, you need to use a join.

Join your data

When you join data, the files must have at least one field in common. Your files share the **Order ID** and **Product ID** fields, so you can join on those fields to see all the rows that have those fields in common. You remember an option to create a join when you created your union using drag and drop, so you give that a try.

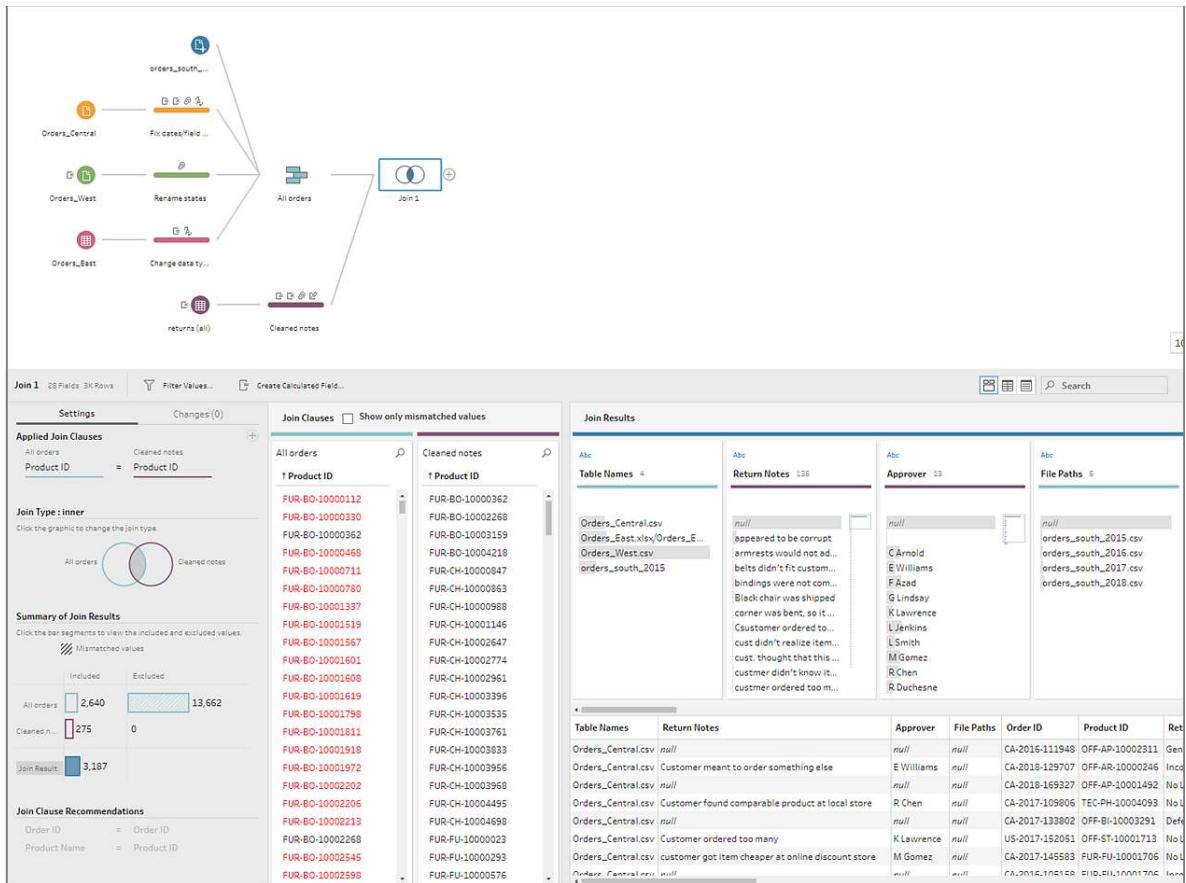
1. In the **Flow** pane, drag the **Cleaned notes** step on to the **All orders** Union step and drop it on **Join**.



When you join files, Tableau Prep shows you the results of your join in the **Join Profile**.

Working with joins can be tricky. You often want to have a clear view of the factors that are included in the join, such as the fields used to join the files, the number of rows included in the results and any fields that aren't included or are null values.

As you review the results of the join in Tableau Prep, you are delighted to see so much information and interactivity at your fingertips.



Tip: The far left pane of the join profile is where you can explore and interact with your join. You can also edit values directly in the **Join Clauses** panes and perform cleaning operations in the **Join Results** pane.

Click in the **Join Type** diagram to try different join configurations and see the number of rows included or excluded in your join for each table in the **Summary of Join Results** section.

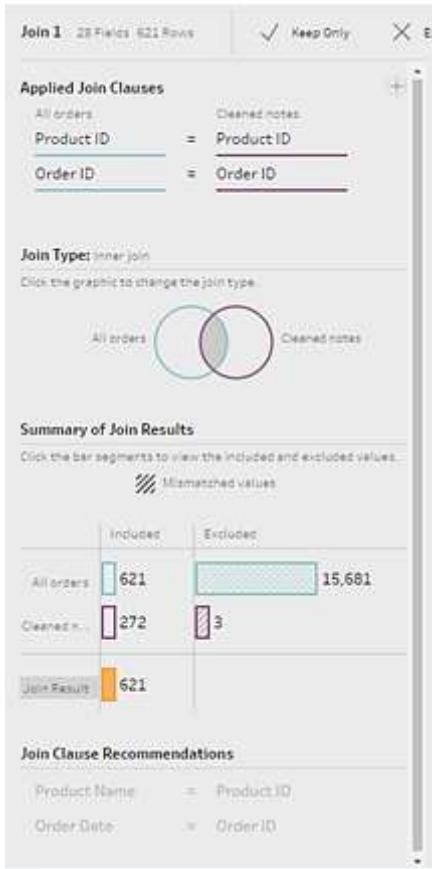
Select the fields that you want to join on in the Applied Join Clauses section or add suggested join clauses from the Join Clause Recommendations section.

For more information about working with joins, see [Aggregate, Join, or Union Data](#).

You see that you have over 13,000 rows excluded from your **All Orders** files. When you created your join, Tableau Prep automatically joined on the **Product ID** field, but you also wanted to join on the **Order ID** field.

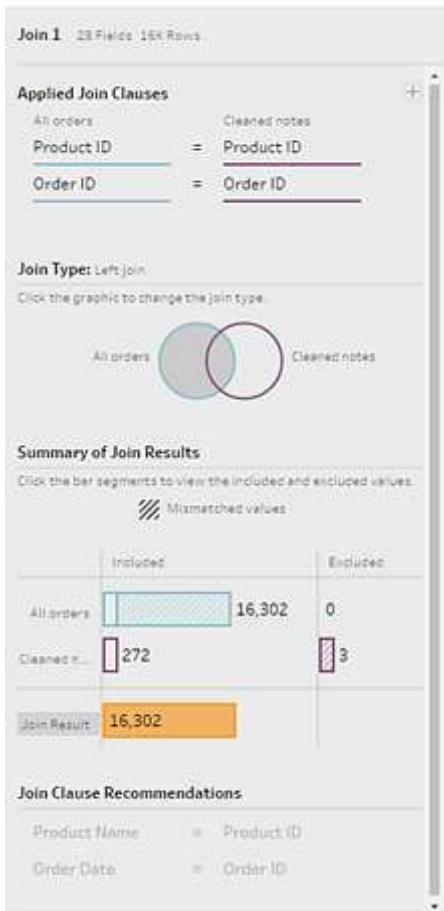
As you scan the left pane of the join profile, you see that **Order ID** is in the list of recommended join clauses, so you quickly add it from there.

2. In the left pane of the **Join** profile, in the **Join Clause Recommendations** section, select **Order ID = Order ID** and click the plus  button to add the join clause.



Because the **Join Type** is set to an inner join (the default setting for Tableau Prep), the join is only including values that exist in both files. But you want all of the data from your **Orders** files as well as the return data for those files. So you'll need to change the join type.

3. In the **Join Type** section, click the side of the diagram to include all orders. In the example below, click the left side of the diagram to change the join type to a **Left** join and include all data from the **All orders** union step and any matching data from the **Cleaned notes** step.



Now you have all of the data from the sales order files and any return data that apply to those orders. You review the **Join Clauses** pane and see the distinct values that don't exist in the other file.

For example there are many order rows (shown in red) that have no corresponding return data. You love being able to explore this level of detail about your join.

You're anxious to start analyzing this data in Tableau Desktop, but you notice a few results from the join that you want to clean up before you do that. Good thing you know what to do!

Tip: Wonder if your data is clean enough? From Tableau Prep Builder, you can preview your data in Tableau Desktop from any step in your flow to check it out.

Just right-click on the step in the **Flow** pane and select **Preview in Tableau Desktop** from the menu.

You can experiment with your data and any changes that you make in Tableau Desktop won't write back to your data source in Tableau Prep Builder. For more information see [View flow output in Tableau Desktop](#)

4. Before you start cleaning your join results, name your **Join** step **Orders+Returns** and save your flow.

Clean your join results

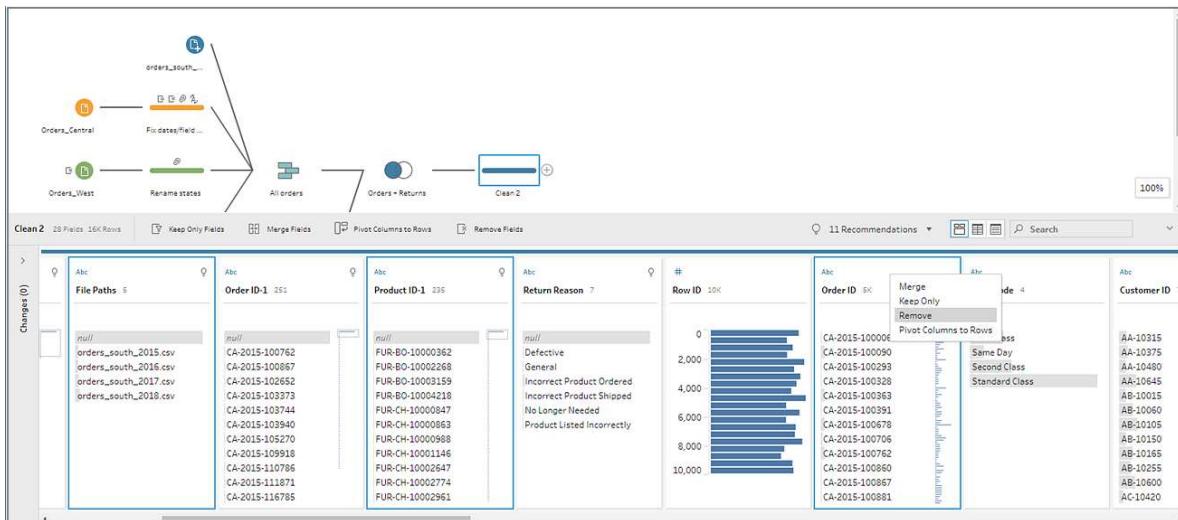
Note: To clean up the fields in your join, you can perform cleaning operations directly in the Join step. For the purposes of this tutorial we will add a cleaning step so you can clearly see your cleaning operations. If you want to try performing these steps directly in the join step skip steps 1 and 3 below.

When you joined the two steps, the common fields Order ID and Product ID were added for both tables.

You want to keep the Product ID field from all of your orders and the Order ID field from the returns file and remove the duplicate fields that came from those files. You also don't need the **File Paths** and **Table Names** fields in your output file, so you want to remove those fields as well.

Tip: When you join tables using fields that exist in both files, Tableau Prep brings in both fields and renames the duplicate field from the second file by adding a "-1" or a "-2" to the field name. For example Order ID and Order ID-1.

1. In the **Flow** pane, select **Orders+Returns**, click the plus  icon, and add a clean step.
2. In the **Profile** pane, select and remove the following fields:
 - **Table Names**
 - **Order ID**
 - **File Paths** (Tableau Prep Builder only)
 - **Product ID-1**



3. Rename the field **Order ID-1** to **Order ID**.

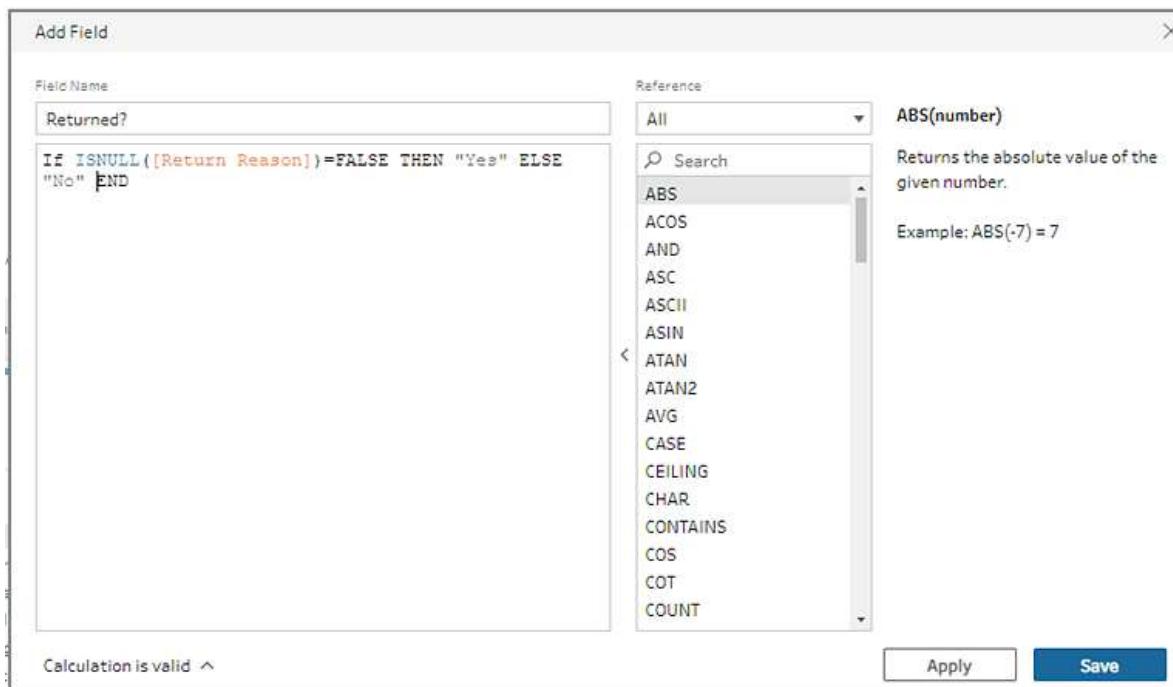
You have quite a few null values where the product was returned but there was no return note or approver indicated. To make this data easier to analyze, you want to add a field with a value of **Yes** and **No** to indicate whether the product was returned.

You don't have this field, but you can add it by creating a calculated field.

4. In the toolbar, click **Create Calculated Field**.

5. Name the field **Returned?** and then enter the following calculation and click **Save**.

```
If ISNULL([Return Reason])=FALSE THEN "Yes" ELSE "No"  
END
```

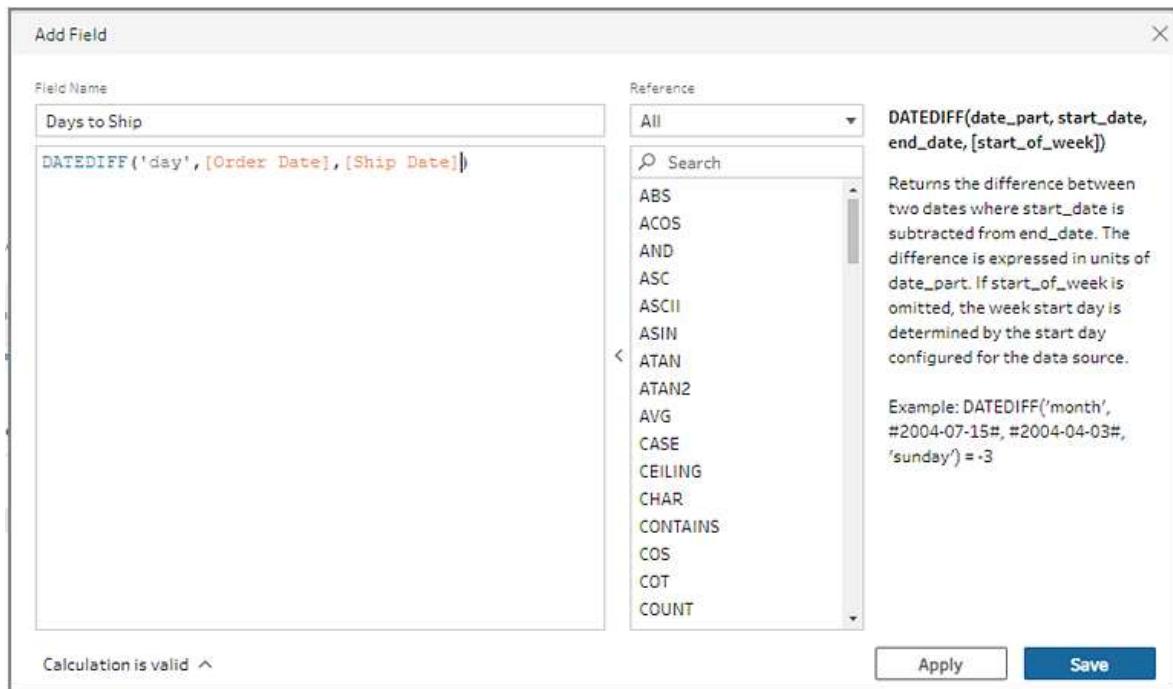


For your analysis you would also like to know the number of days it takes to ship an order, but you don't have that field either.

You have all the information that you need to create it though, so you add another calculated field to create it.

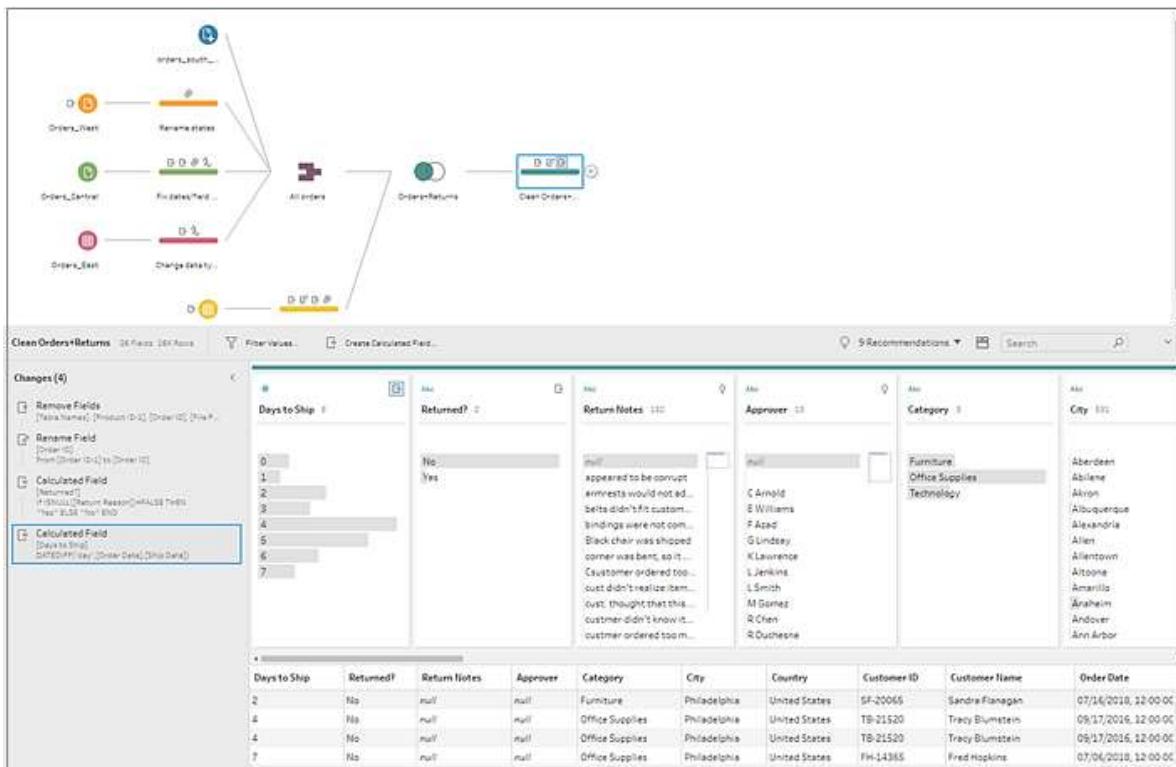
6. In the toolbar, click **Create Calculated Field**.
7. Name the field **Days to Ship** and then enter the following calculation and click **Save**.

`DATEDIFF('day', [Order Date], [Ship Date])`



8. Name your step **Clean Orders + Returns**.

9. Save your flow.



5. Run your flow and generate output

Your data is looking good and you're ready to generate your output file to start analyzing it in Tableau Desktop. All you need to do is run your flow and generate your extract file. To do this you need to add an **Output** step.

Depending on where you're working, you can output your flow to a file (Tableau Prep Builder only), to a published data source or to a database.

1. In the **Flow pane**, select **Clean Orders+Returns**, click the plus icon and select **Output (Add Output in prior versions)**.

When you add an Output step, the **Output** pane opens and shows you a snapshot of your data. Here you can select the type of output that you want to generate, and specify the name and where you want to save the file.

The default location is in the **My Tableau Prep Builder** repository in your data sources folder.

The screenshot shows the Tableau Prep Builder interface. On the left, the **Flow pane** displays a data pipeline starting with three input sources: **Orders_Central**, **Orders_West**, and **Orders_East**. These sources undergo various transformations: **Fix dates/field...**, **Rename states**, and **Change data type...**. The transformed data then flows through an **All orders** step, followed by an **Orders + Returns** step, and finally a **Clean Orders + ...** step. The final output is directed to the **Output** step. On the right, the **Output** pane is open, showing the following configuration:

- Save output to:** File
- Name:** Output
- Location:** C:\Data\Datasources
- Output type:** Tableau Data Extract (.hyper)
- Write Options:** Select an option to create or update your output table.
- Full refresh:** Create table

The pane also displays a preview of the data with 26 fields, including columns like Days to Ship, Returned?, Return Notes, Approver, Order ID, Return Reason, Row ID, Ship Mode, Customer ID, Customer Name, Segment, Country, City, State, Postal Code, and Product ID. The data preview shows several rows of customer information from different regions and product categories.

2. In the left pane in the **Save output to** drop-down, depending on where you are working, do one of the following:

Tableau Prep Builder

- Select **File** (select **Save to file** in previous versions).
- Click the **Browse** button, then in the **Save Extract As** dialog, enter a name for the file, for example **Orders_Returns_Superstore**, and click **Accept**.
- In the **Output type** field, select an output type. Select Tableau Data Extract (.hyper) for Tableau Desktop or Comma Separated Values (.csv) if you want to share the extract with a third party.

Tableau Server or Tableau Online

- Select **Published data source**.
- Select a project.
- Enter a name for the file, for example **Orders_Returns_Superstore**.

Tip: You have choices when generating output from your flow. You can generate an extract file (Tableau Prep Builder only), you can publish your data as a data source to Tableau Server or Tableau Online or you can write your data to a database. For more information about generating output files, see [Create data extract files and published data sources](#).

3. In the **Write Options** section, view the options to write the new data to your files. You want to use the default (**Create table**) and replace the table with your flow output, so there is nothing to change here.

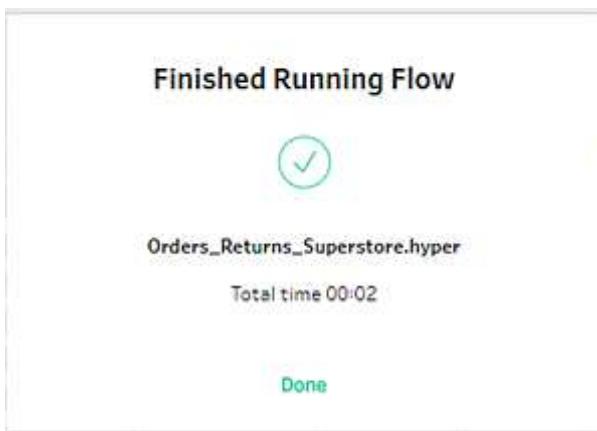
Tip: Starting in version 2020.2.1, you can choose how you want to write your flow data back to your table. You can choose from two options; **Create table** or **Append table**. By default, Tableau Prep uses the **Create table** option and overwrites your table data with the new data when you run your flow. If you choose **Append table**, Tableau Prep adds the flow data to the existing table so you can track both new and historical

data on every flow run. For more information, see [Configure write options](#) ([prep_incremental_refresh.htm#flow-writeoptions](#)).

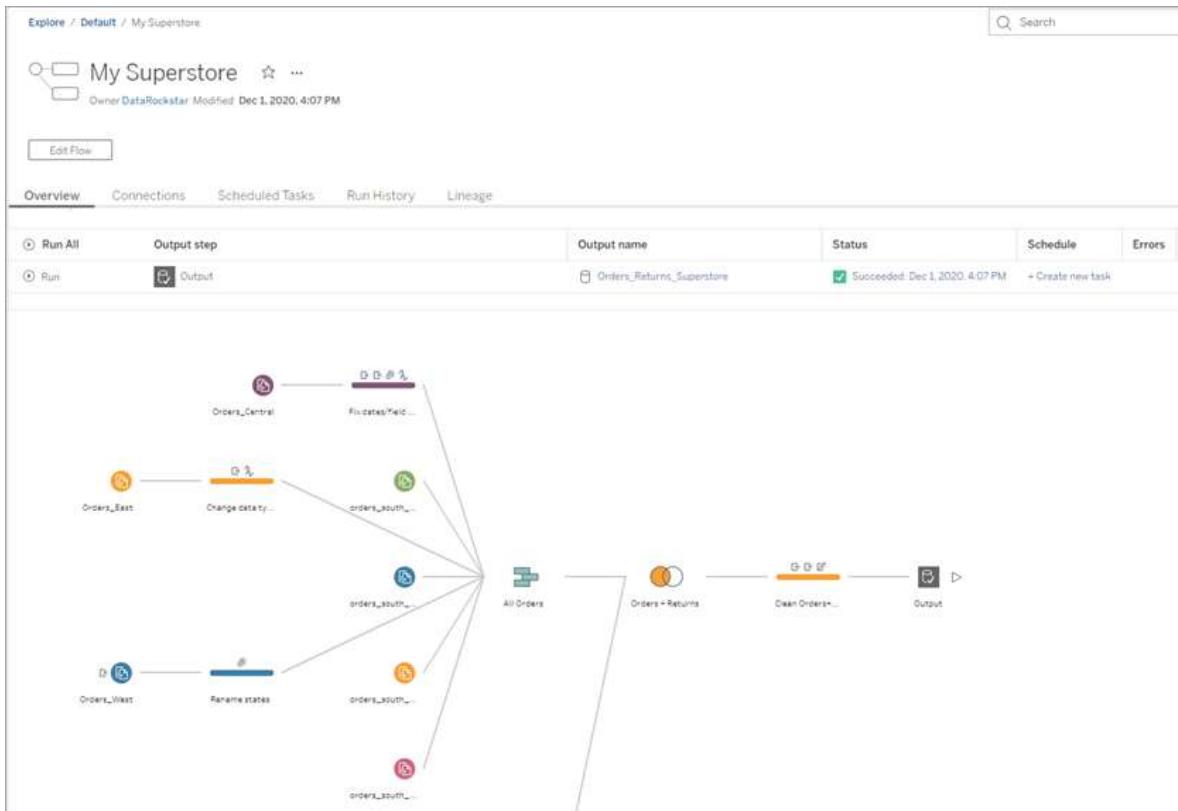
4. In the Output pane, click **Run Flow** or click the **Run Flow** ▶ button in the flow pane to generate your output.

Note: If you are working on the web, click **Publish** to publish your draft flow. Only published flows can be run.

5. When the flow is finished running, a status dialog shows whether the flow ran successfully and the time it took to run. Click **Done** to close the dialog.



If working on the web, navigate to the **Explore>All Flows** page, and find your flow. You can see the status of your flow run on the Flow Overview page.



To keep your data fresh, you can run the flow manually or use the command line. If you have the Data Management Add-on and have Tableau Prep Conductor enabled, you can also run your flow on a schedule in Tableau Server or Tableau Online.

Starting in Tableau Prep Builder version 2020.2.1 and on the web, you can also choose to refresh all your data every time the flow is run, or run your flow using incremental refresh and process only your new data each time.

For more information about keeping your data fresh, see the following topics:

- [Refresh flow output files from the command line](#)
- [Publish a Flow to Tableau Server or Tableau Online](#)
- [Refresh Flow Data Using Incremental Refresh](#)

Wrap up and resources

You are a data prep rock star! You took dirty data and transformed it with ease! In no time, you cleaned and prepped your data from multiple data sets and turned it into a sleek, clean data set that you can now work with in Tableau Desktop to do your analysis.

Want more practice? Try replicating the rest of the sample flow for Superstore using the data files found here:

- Orders South 2015
(https://mkt.tableau.com/PrepTutorials/orders_south_2015.csv).
- Orders South 2016
(https://mkt.tableau.com/PrepTutorials/orders_south_2016.csv).
- Orders South 2017
(https://mkt.tableau.com/PrepTutorials/orders_south_2017.csv).
- Orders South 2018
(https://mkt.tableau.com/PrepTutorials/orders_south_2018.csv).
- Orders Central
(https://mkt.tableau.com/PrepTutorials/Orders_Central.csv).
- Orders East
(https://mkt.tableau.com/PrepTutorials/Orders_East.xlsx).
- Orders West
(https://mkt.tableau.com/PrepTutorials/Orders_West.csv).
- returns reasons new
(https://mkt.tableau.com/PrepTutorials/return_reasons_new.xlsx).
- Quota (<https://mkt.tableau.com/PrepTutorials/Quota.xlsx>).

You can also find the files in the following location on your computer after installing Tableau Prep Builder:

- (Windows) C:\Program Files\Tableau\Tableau Prep Builder <version>\help\Samples\en_US\Superstore Files
- (Mac) /Applications/Tableau Prep Builder <version>.app/Contents/help/Samples/en_US/Superstore Files

Want more training? Check out the [training videos](#) for Tableau Prep Builder, check out these great [resources](#), or take an [in-person training](#) course.

Want more information about the topics we covered? Check out the other topics in the Tableau Prep online help.

Other articles in this section

- [About Tableau Prep \(prep_about.htm\)](#).
- [Tableau Prep Visual Dictionary \(prep_visual_dictionary.htm\)](#).
- [About Tableau Help \(about-tableau-help.htm\)](#).